DOCUMENT RESUME

ED 350 111 PS 020 880

AUTHOR	Fontaine, Denis
TITLE	Evaluating Health Action in the Third World.
INSTITUTION	International Children's Centre, Paris (France).
REPORT NO	ISSN-0379-2269
PUB DATE	92
NOTE	72p.
AVAILABLE FROM	International Children's Centre, Chateau de Longchamp, Bois de Boulogne, 75016 Paris, France ($10; 6-issue annual subscription $40).
PUB TYPE	Guides - Non-Classroom Use (055) -- Collected Works - Serials (022)
JOURNAL CIT	Children in the Tropics; n198 1992
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Community Health Services; Developing Nations; *Evaluation Methods; Foreign Countries; Health Personnel; *Health Programs; Medical Care Evaluation; Mortality Rate; Nutrition; Professional Training; Program Effectiveness; *Program Evaluation; Program Implementation; *Public Health
IDENTIFIERS	Morbidity; Program Objectives; Service Utilization

ABSTRACT

The theme of this serial issue is helping individuals involved in community health programs at the local level to devise and apply a protocol for the evaluation of a community health program. In the opening sections of the issue evaluation is defined, the difference between a direct quantifying measure and an indicator is clarified, obstacles to accurate evaluation are identified, and steps in an evaluation process are outlined. This is followed by a section concerned with evaluation of program action, which addresses the financial, material, and human resources of a program; the amount of planned activity that was accomplished; and program implementation variables, such as the population that was served by the program. Indicators which are particularly useful in evaluating health programs are also discussed. The next section, on evaluation of program effects, addresses program results, or which specific objectives were met; the overall program impact upon the target population; and the strategy used by the program to implement its goals. Two other sections discuss indicators useful in evaluating health programs (these include availability, accessibility, utilization, and quality of service); and evaluation of the training and supervision of health service personnel. A final section dealing with evaluation of the health status of a population, points out that surveys of indicators of morality, morbidity, and nutrition are required. Interspersed throughout the issue are six self-evaluation exercises, answers to which are provided at the end of the document. A 15-item list of English and French publications for further reading is provided. (BC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

30th year

CHILDREN IN THE TROPICS

ED350111

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

EVALUATING HEALTH ACTION IN THE THIRD WORLD

1992 - N° 198

PS C20880

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

P. Vesin

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

INTERNATIONAL CHILDREN'S CENTRE – PARIS

The International Children's Centre was created by the French government in 1949, on the initiative of Professor Robert Debré in particular, following negociations between France and the United Nations. Its purpose was to furnish those international and national agencies dealing specifically with child care with training facilities and educational and informational tools in the field of child health and development, viewing children within their family and surroundings.

ICC soon turned essentially toward Third World children and devoted its activities to the training and education of personnel with social, educational and administrative responsabilities as well as medical and paramedical workers. The desire for greater efficiency has led it to work increasingly with trainers and to concentrate its efforts on the methodological and educational aspects of mother and child care programmes.

ICC is also engaged in an attempt to further study — and — action on some aspects of the life and health of children and their family, so as to contribute to practical improvement, particularly in the fields of growth, nutrition, planned parenthood, the control of transmissible and nutritional diseases, preschool and school education, the needs of disabled and underprivileged children, etc.

Over this period of more than 30 years, a large amount of documents on children and adolescents, mostly from the developing countries, has been accumulated. This international documentation has been classified and sorted out, and has been computerized since 1983 : a bibliographic data base (BIRD : « Base d'Informations Robert Debré ») may be consulted anywhere in the world, through international communications networks. ICC also publishes periodicals, educational documents and specialized bibliographic bulletins.

As for its legal status, the International Children's Centre is a foundation under french law of recognized public utility, administered by an executive board with broad international membership.

/

BEST COPY AVAILABLE

CHILDREN
IN THE TROPICS

REVIEW OF THE INTERNATIONAL CHILDREN'S CENTRE

EVALUATING HEALTH ACTION IN THE THIRD WORLD

☐ DENIS FONTAINE

1992 - N° 198

THE FRENCH ASSOCIATION OF VOLUNTEERS FOR PROGRESS

The objective of the French Association of Volunteers for Progress (FAVP) is to contribute to the development of those countries viewed as most underprivileged.

To do so, the FAVP is continually involved in 300 development operations in a wide range of fields. At present, more than 550 young people, both boys and girls, are working on such projects hand in hand with the local population, and for its benefit, in 30 African countries.

These volunteers for progress spend two years acquiring a most enriching professional experience, along with a quite unique human experience.

The FAVP participates in health development operations with two main goals :

- improving the health status of population groups : our work is of the public health type, as opposed to action aimed at individual health. In collaboration with the health authorities, the following are planned :

. action over an extended geographic area, a district or at most, a region. The size of the target population ranges from 10,000 to 300,000 inhabitants ;

. priority is given to primary health care, with emphasis on maternal and child health in conjunction with nurses and village health agents.

- reinforcing the autonomy of the local population and health system. Most projects aim at reinforcing the public health system of the countries in which we operate.

Several themes have received special attention in recent years :

- the reinforcement and enhanced autonomy of health systems through the establishment of partial recovery of running costs. The main objective is the strengthening of basic health systems, so that they may offer high quality comprehensive services (prevention and care) on an on-going basis ;

- maternal and child health programmes : programmes for immunization, nutrition, anti-dehydration, surveillance of pregnancies, etc ;

- Aids prevention programmes, centred around groups with high-risk behaviour patterns.

FAVP is currently engaged in 40 health-related operations in 16 African countries : Benin, Burkina Faso, Burundi, Cameroon, Central African Republic, Congo, Djibouti, Gabon, Guinea, Guinea Bissau, Ivory Coast, Mauritania, Niger, Ruanda, Senegal and Togo. 70 volunteers, mostly nurses, are involved in the implementation of these programmes.

Denis Fontaine, community health physician, has been in charge of FAVP health programmes since 1990. He worked as a volunteer in the Congo from 1986 to 1988.

SECRETARIAL WORK SYLVIANE LE BIHAN

TRANSLATION HELEN ARNOLD

EVALUATING HEALTH ACTION IN THE THIRD WORLD

TO EVALUATE OR NOT TO EVALUATE?		5
EVALUATION, TRUTHS AND FALSE-HOODS	Definition	9
	General method	9
	"False evaluations"	12
	Who performs evaluation?	13
EVALUATION, EVALUATIONS	Definitions	15
	Steps in planning	16
EVALUATING ACTION	Evaluation of resources	21
	Evaluation of activities	22
	Evaluation of implementation	24
	Choice of indicators	27
	Decisions made following evaluation	30
	Limits of action-centred evaluation	30
EVALUATING EFFECTS	Evaluation of results	32
	Evaluation of impact	35
	Evaluation of strategy	35
SUCCESSFUL EVALUATION	Phases in evaluation	37
EVALUATING A HEALTH UNIT	Availability	39
	Accessibility	41
	Utilization	41
	Adequate utilization	42
	The coverage diagram	43
EVALUATING TRAINING AND SUPERVISION OF PERSONNEL	Training	45
	Supervision	50

EVALUATING THE HEALTH STATUS OF A POPULATION
Indicators of mortality 52
Indicators of morbidity 54
Nutrition-related indicators 56

WHAT IS TO BE EXPECTED OF EVALUATION ? 59

ANSWERS TO THE SELF-EVALUATION EXERCISES 61

FURTHER READING... 65

TO EVALUATE OR NOT TO EVALUATE?

One day, a man decided to empty the ocean. He had many reasons to want to empty the ocean, but there was one most important reason : there is not enough room on land for everyone. So he took a pail and began to empty the ocean. A year later, another man came along.

- What are you doing?

- I'm emptying the ocean, can't you see?

- With a pail?

- Well, it's better than using a tin can !

"You're tiring yourself out for nothing", concluded the second man, and he shrugged his shoulders and went away.

The man was alone with his pail again. He was irked that his efforts had not been fairly evaluated. He sat down for a while, thought about the situation, then bought a pen and a notebook and returned to his pail. Another year went by, and another man came along.

- What are you doing?

- I'm emptying the ocean, can't you see?

- With a pail?

- Absolutely, with a pail !

He took out his notebook, and the most seriously in the world, said : "Look, I've worked twelve hours a day every day for one year, and I've emptied two hundred twenty thousand pails. That makes three million five hundred thousand liters of water !"

He raised his head to see the effect these figures produced on his listener. But instead of seeing him serious and admiring, he saw only an ironic smile.

- That's good ! And...has the level of the ocean gone down?

"Well...surely !" he answered, surprised.

But the other man looked at the ocean and shrugged his shoulders : "You bet, it's low tide !" he concluded.

The emptier of the ocean was very angry. For the second time, his work has not been fairly evaluated. He went back to his pail, but less enthusiastically. Sometimes he found himself staring out at the ocean. In fact, a question was bothering him : had the level of the ocean really gone down? It seemed to him that it had ; something inside him said it truly had : he really had removed those millions of liters ! But how could he be sure, when the tide came in and out all the time?

At other times, when the weather was grim and the wind blew, or when a higher tide than usual came in, he had doubts. And then he thought : the ocean will never be emptied if he stood around doing nothing. After all, it is a long-term job : when he will have done more the results will be obvious for everyone, and he will no longer have to justify himself. And then, he thought, people are glad I can provide them with salt ...

So he returned to his pail and continued to empty the ocean.

How will this story end? There are three probable solutions :

- the man will go on with his project, and end his life exhausted, with a pail in hand. One variation : sapped by the criticisms of passersby, he will lose his mind before he can achieve his goal ;

- one day, or one night, he will trample his pail and go off to a city to make a new life for himself, and will curse generous idealists for the rest of his life ;

- he will turn to producing salt.

Now let us look at the situation with the eyes of an evaluator. What does he/she see? First, that the emptier of the ocean had a noble goal, but did not determine whether his programme was realistic in respect to his ambition. In this particular case, a prior evaluation would certainly have shown that a single pair of arms plus a pail is an inadequate means of emptying an ocean.

The action of the emptier of the ocean may have been courageous, but he launched upon it without making the effort to evaluate it. As a result, the first passerby could make fun of him. He learned from this experience, and attempted an evaluation : by noting the work done daily, he was able to evaluate the resources employed (the time spent working) and the process (his activity, or the number of pails or liters of water removed). This is praiseworthy in itself, and in terms of figures, he may be said to have worked "well".

But then another skeptic came along, and showed not the slightest interest in this type of evaluation, since he was interested in concrete results. The emptier, who was completely absorbed in his work, had completely overlooked this aspect of the problem. Intuitively, he thought that there was necessarily a result, since there was an action as planned. But the passerby disagreed, and considered that the "measurement" was distorted by the low tide.

The emptier of the ocean was also disturbed by this difficulty, which he had not foreseen, and his morale ebbed and flowed with the tides. He then sought consolation in the impact of his action : he provided people with salt. If he had used some statistical tools, he would have been able to overcome this measurement problem, and discover that the mean level of the ocean had not been significantly modified. Perhaps he would have rethought his entire plan of action.

Following some such evaluation, what would the emptier of the ocean have done? There seem to be three possible decisions :

- retain the same objective but make use of other resources and/or other types of action such as purchasing a motor pump (change in resources) or bore a hole in the bottom of the ocean so that it will drain out by itself (change in type of action), with the understanding that each postulate should be evaluated ;

- choose another more realistic objective for the same activity, such as producing salt ;

- interrupt the programme and change activities : for instance, practice market gardening, in which he could put his experience to use when watering his garden.

Our initial question remains : is evaluation necessary or not? Did the emptier of the ocean ever do an evaluation? Actually, the answer is yes. Intuitively, almost subconsciously, he evaluated every phase in his programme : he started the action because he felt it was possible, he was able to extract "large" amounts of water and the level of the ocean seemed to drop, but with the results described above.

The choice is not between evaluating and not evaluating, but between intuitive evaluation based on a few impressions, and strict evaluation based on systematic data-collection covering all areas within the programme. A comparison of the results of these two types of evaluation shows that their conclusions may sometimes coincide (in the present case, in reconversion to producing salt). However, strict evaluation has one great advantage over intuition : it avoids exhausting oneself at useless work. This example may be eloquent for a nurse or a doctor disposing of little means, and engaged in an ambitious, demanding action but one whose effects are hard to measure, as may be the case in a community health programme.

The present paper is designed as an aid in self-instruction. It was written for people with responsibilities in community health programmes at the local (village or district) level, to the exclusion of emergency aid.

Evaluation of planning, economic evaluation and the evaluation of research are not discussed in detail.

The objectives of this guidebook are as follows.

GENERAL OBJECTIVE

After reading this paper, and with the help of the tools provided, the reader should be able to devise and apply a protocol for the evaluation of a community health programme.

SPECIFIC OBJECTIVES

The reader should be able :

- to define evaluation and oppose it to "false" evaluations ;

- to define the different types of evaluation and to class them according to the level involved (planning, action, effects) ;

- to describe the method of evaluation at the action level (resources, activities) and with respect to effects (results, impact) ;

- to list the information required for each type of evaluation ;

- to select appropriate indicators and criteria for the type of action conducted and the level of evaluation desired ;

- to list prerequisites for successful evaluation.

Practice in self-evaluation is proposed at the end of each chapter.

EVALUATION, TRUTHS AND FALSEHOODS

Evaluation is a value judgement grounded in objectivity, but attention must be paid to the psychological context.

DEFINITION

To evaluate is to determine the value of something, to make a value judgement. More specifically, in the case of community health, it involves assessing a programme of activities, and we then speak of programme evaluation. Evaluation then involves the measurement of the results of action with respect to its objectives. In other words, to evaluate is to answer the question of the value of what is being done, or more exactly, of the extent to which the desired result has been attained.

Evaluation is not an end in itself ; it is a means of drawing lessons from experience, an instrument aimed at decision-making. Evaluation has two main goals, which are not mutually exclusive :

- by the constructive criticism of action and of its consequences, evaluation is useful in improving a programme, and in correcting the errors committed in the course of an intervention. In a certain sense, it is a learning situation : following action, evaluation sorts out what was positive and should be continued from what did not work and should be modified. This is why we speak of formative evaluation ;

- the approach may be more comprehensive, and the questions then pertain to whether the programme should be stopped, continued or extended. An overall appraisal is made, and the effects of the programme are measured, including those defined in the original objectives and those that are more or less unexpected. We then speak of summative evaluation.

Evaluation is not an isolated process, but one that is a part of the "programme management" cycle ; on the basis of health problems viewed as requiring priority, an action programme aimed at solving them or at least attenuating them is devised. This programme is implemented, then evaluated. Evaluation results in a different vision of health problems, and the programme is then modified, and so on. This is a spiralling process, since some progress is made each time the cycle comes full round.

It is understandable, then, that evaluation must be foreseen and incorporated at the planning stage. And yet, health action is often launched on a more or less intuitive basis, with no accurate planning : even then, evaluation remains feasible and becomes the first step in planning for the next cycle.

GENERAL METHOD

How is this value judgement known as evaluation made? Since a high level of objectivity is desired, quantitative assessment - that is, measurement - is used whenever possible. Some types of measurement such as a purchasing price or the number of patients

seen at a consultation are easily obtainable. Other elements are not directly measurable, or even observable, however : this is the case of the state of health of pregnant women or the community's participation in health action. An effort must then be made to delimit the subject using its observable aspects. For instance, the health status of pregnant women may be approached through the maternal mortality and morbidity rates. Community participation may be appraised by the observation of community achievements such as a collective garden or various buildings. This work is normally done during the planning phase, or when objectives (the expected results of action) are being defined.

Indicators

Specific objectives defined by this process may be directly quantified. Maternal mortality, for instance, is measured by the maternal mortality rate (the ratio of the number of deaths of pregnant women or of mothers within six weeks of childbirth to the number of births, over a one-year period).

When specific objectives cannot be assessed, variables that measure them indirectly are sought : these are known as indicators because they provide an indication about the fact studied. For example, the morbidity of pregnant women cannot be measured directly, since there is no way of comparing anaemia, so frequent and often unnoticed, with eclampsia, which is extremely uncommon but very serious. Indicators are therefore needed. In a heavily malaria-ridden area, the number of episodes of malaria during pregnancy may be an indicator of maternal morbidity. The percentage of dystocic deliveries is an indicator of maternal morbidity at the time of delivery.

Single indicators only rarely describe a phenomenon correctly, but it is not advisable to use too many at once. Data collection would then become too burdensome and probably be poorly performed. It is therefore important to choose few indicators, but good ones, the main qualities of which are :

- validity : an indicator is valid if it really measures the fact studied. This is the main quality of any indicator. For instance, the average number of antenatal consultations per pregnant woman is not a valid indicator of maternal morbidity, since regular attendance of health services is not automatically a sign of good health. Conversely, it is a valid indicator of use of maternal and child health services (MCH).

- objectivity : an indicator is objective or reliable if the same measurement done by different persons arrives at the same results ;

- availability : an indicator is available if the data is easy to collect, at a tolerable cost and in a way that is acceptable to the community ;

- sensitivity : an indicator is sensitive if it changes as soon as the fact observed changes. This is important, since evaluation is an attempt to detect change ;

- specificity : an indicator is specific of a fact if it measures nothing but that fact. This is a more infrequent, less important quality than sensitivity. Low specificity should incite caution in interpretation, since a different figure for an indicator may be due to a cause other than a change in the fact studied.

Measurement of variables related to the objectives and indicators does not suffice ; judgement must be based on a comparison of this measurement with a reference value, called a criterion or norm. This reference value may :

- be determined by the scientific community, as is the case for the weight-for-age of small children, or the cut-off values for high blood pressure ;

- be set by the ministry of health, as when four antenatal consultations are prescribed for pregnant women, or an immunization schedule determines the age at which children should receive each injection ;

- be recommended in the framework of programme objectives such as the reduction of maternal mortality to five per thousand in 5 years, or a check on the immunization record of all children under 3 seen in consultation.

This reference value is not necessarily determined from the outset, since the situation may not be sufficiently known yet. The evaluation team decides whether the result is satisfactory or not by setting the reference value afterward. We then speak of implicit norms, whereas in the previous cases the norms were explicit.

Last, there may be no utilizable indicator capable of accounting for certain facts or situations. Rather than eliminating these, a qualitative judgement is then required. All told, evaluation involves measurements enabling a quantitative judgement with reference to norms and criteria, and a qualitative judgement based on non-quantifiable elements.

EVALUATION = MEASUREMENTS AND QUANTITATIVE JUDGEMENT + QUALITATIVE JUDGEMENT.

Psychological difficulties

Seen in this way, evaluation seems to be a purely logical, rational operation, and yet some people are afraid of evaluation and have extremely emotional reactions to it. There are several reasons for this. Evaluation brings to mind mathematics and strict, demanding protocols - in other words, a whole series of more or less esoteric, time and resource-consuming rituals when both are in short supply - and this leads to rejection. The argument of overworking is not tenable, since it is the equivalent of demanding that a car be driven at night without any headlights to economize the battery. Furthermore, routine registers already in existence are often utilizable for the evaluation of activities. However, there is truth in the fact that a modicum of methodology and knowledge of statistics is necessary.

More seriously, evaluation is anxiety-producing because it may point out some negative and unsuccessful aspects of programmes, which are often painful for those in charge of it. There may be two types of attitudes behind this anxiety. First, there may be the desire to forget a setback ; this may be unimportant if it applies to past events, and much may be learned by understanding the reason behind a failure. On the other hand, if the conflict is still acute, it is preferable to delay evaluation to avoid interference. The other reaction is the fear of being accused of certain errors. This attitude explains why evaluation is always viewed with some suspicion. To minimize this tendancy, it is preferable to talk about it openly, and to specify that no individual will be attacked. It is out of the question to put errors and failures aside, but a distinction must be made between acts, which must be questioned, and individuals, whose right to error must be acknowledged.

The notion of evaluation is somewhat fuzzy, and as a result it may be given a different meaning by each person, the outcome being a distorting of the evaluation process. We must therefore have a clear idea of what evaluation is and is not : evaluation is not a control nor is it make-believe, self-flogging or programme follow-up.

"FALSE EVALUATIONS"

Evaluation cannot be assimilated to a control, which is a more or less bureaucratic procedure aimed at determining right and wrong, and involving the notion of punishment. Evaluation does include a critical aspect, but one which is totally different in spirit and form : its criticism is always constructive and never aims at putting blame, but at understanding why and how an action did not produce all of the expected results. Evaluation attempts to understand things from the inside, to promote change, to integrate new elements, whereas control is limited to the application of a rigid, external norm which cannot be questioned. Evaluation is a "democratic" process in which each participant has something to say, whereas control, which is sent down by a higher authority, is authoritarian and hierarchic in nature.

There is another, more discrete form of control which participates in a technocratic approach : it consists of only viewing reality through the objectives of the programme, which must be achieved at all costs. The head of the programme may then unwittingly take advantage of evaluation to distribute good, and mostly bad marks to the protagonists susceptible of jeopardizing "his/her" objectives. Objectives are a guide for action, and the programme must be adjusted to the situation, and not the reverse.

It must be admitted that no-one is devoid of the temptation to exert control, which is rooted in everyone's mentality. Some precautions must therefore be taken to avoid transforming evaluation into a drama that incites participants to sabotage it. From the outset, it is a good idea to review the true objectives of evaluation and to make it clear, by one's own attitude, that these are really the only ones. All participants in the programme, from the data-collec-

tion point to conclusion-drawing, should be included. The evaluator too must give up the dream of "knowing everything about everything" and accept some approximations and doubts.

Evaluation is not make-believe : it may be side-tracked from its explicit course, and pursue unmentioned and sometimes unmentionable objectives, such as advertising, where the programme is a "product" to be sold to donors or a means of self-aggrandizement. It may also be a camouflage operation, aimed at hiding some unsavoury aspects through apparently positive information, or a diversion, so as to delay true evaluation by performing a symbolic act or side-tracking attention to minor problems. One must also be wary of manipulatory operations aimed at attacking individuals and discrediting them under the pretext of evaluation.

All of these pseudo-evaluations are based on the idea (sometimes true) that what is written in a report actually exists, and may be more real than reality itself ! Once again, if some aspects of a programme require criticism they should be criticized, but it is action that should be challenged and not individuals.

Nor is evaluation a self-flogging operation : such pseudo-evaluation aims at depreciating the operation by complacently emphasizing failures and negative aspects, while eluding or denying the positive results. This attitude is seen in depressive evaluators who no longer believe in their work, and the outcome is discouragement among all participants.

Evaluation-self-flogging may very well be based on actual facts, but it is nonetheless a pseudo-evaluation since it is incomplete, does not respect the respective weight of different events and does not provide any constructive solutions. This attitude may be avoided by the massive participation of protagonists, which diversifies ways of looking at events and revives their real dimensions. It is normal to give more attention to failures and difficulties than to successes, but the positive aspects should also be clearly brought out. Remedies for the failures should always be suggested.

Follow-up is not evaluation. Follow-up and supervision of a programme yield information on the functioning of work such as exchanges of correspondence, encounters with partners, activities reports which provide information on the situation, the advancement of the programme and the difficulties encountered. All of this information is extremely useful, but insufficient. It is the ground on which evaluation, which involves a value judgement, is based. Furthermore, this type of relatively untargetted information may contain considerable gaps pertaining to data which is essential for decision-making. If this information-collecting work is to be used in evaluation, it is preferable that a record be kept of all important decisions, the reasons behind the choices and any occurrences.

WHO PERFORMS EVALUATION?

As a rule, protagonists in the programme should participate in evaluation : they are familiar with the programme and their opinions and the information they continually record must be taken into

account. Further, this straightens out some psychological problems.

Should the evaluation leader be a member of the programme team? If he/she is an outside person, the evaluation is said to be external. This solution is valuable in that it provides a new vision, freed of the thousand and one everyday constraints involved in the programme. This vision, personified by an experienced and therefore credible individual, brings with it a passion-free judgement. However, bringing someone in to do evaluation involves great expense, the person may not know enough about the local context and the participants in the programme may experience the intrusion of an outsider as a threat to themselves.

If the evaluation leader is a member of the team, the evaluation is said to be internal or self-evaluation. In this case, a person who is familiar with the programme is available at a low cost, but his/her judgement may be less independent of the material and interpersonal contingencies of the programme, and less creative. Self-evaluation is often the preferred option at the local level in community health work, mostly for financial reasons.

Self-evaluation exercise n° 1

1. Define programme evaluation.
2. What are the two main goals of programme evaluation?
3. List four broad categories of false evaluation. For each type, give a brief explanation of why these are not evaluations.
4. What, in your opinion, are the prerequisites for successful evaluation?

Answers at the end of this issue.

17

EVALUATION, EVALUATIONS

Before the intervention (planning), during it (action) and after action (effects) : each phase has its own evaluations.

There are three major strategies for action aimed at improving public health : direct health action, health development and economic development.

DEFINITIONS

Direct health action aims at reducing mortality and morbidity (this is the general objective). The idea is to use the complete range of preventive and curative measures to combat disease. Education for health aims at changing behaviour, action on the environment attempts to reduce the natural reservoir of diseases, curative care works at reducing the duration and seriousness of diseases, while the purpose of immunization is to enhance the population's immunity, etc.

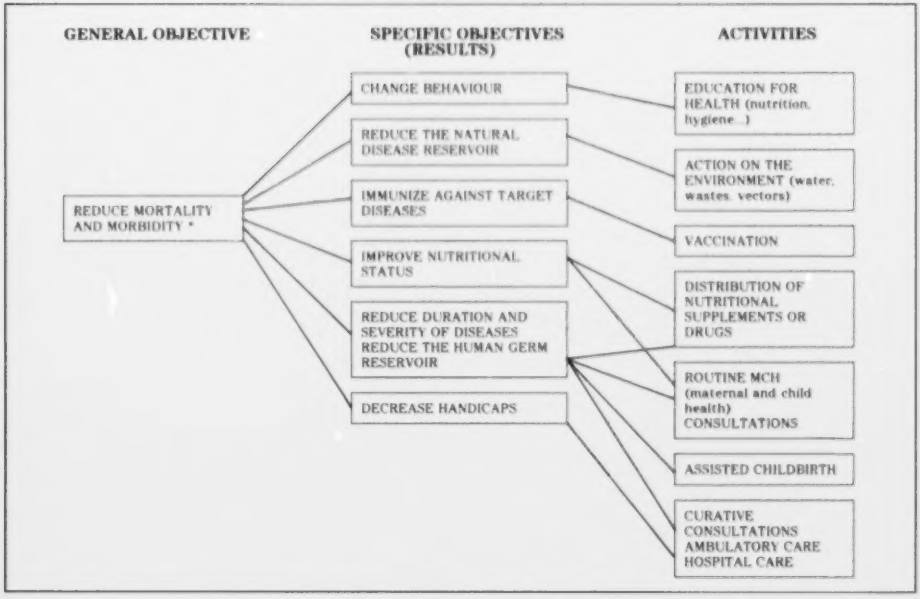

GENERAL OBJECTIVE
SPECIFIC OBJECTIVES (RESULTS)
ACTIVITIES
REDUCE MORTALITY AND MORBIDITY *
CHANGE BEHAVIOUR
REDUCE THE NATURAL DISEASE RESERVOIR
IMMUNIZE AGAINST TARGET DISEASES
IMPROVE NUTRITIONAL STATUS
REDUCE DURATION AND SEVERITY OF DISEASES
REDUCE THE HUMAN GERM RESERVOIR
DECREASE HANDICAPS
EDUCATION FOR HEALTH (nutrition, hygiene...)
ACTION ON THE ENVIRONMENT (water, wastes, vectors)
VACCINATION
DISTRIBUTION OF NUTRITIONAL SUPPLEMENTS OR DRUGS
ROUTINE MCH (maternal and child health) CONSULTATIONS
ASSISTED CHILDBIRTH
CURATIVE CONSULTATIONS AMBULATORY CARE HOSPITAL CARE

Figure 1 : Activities and possible results in direct health action

This figure is read from right to left for evaluation (what result may be expected from these activities?) and from left to right for planning (what activities should be launched to achieve what objective?).

** This general objective may be divided into more specific intermediary objectives based on different population groups (infant mortality, maternal mortality) or diseases (diarrhoeal diseases, measles...).*

Health development is aimed at improving the health system (this is the general objective). This strategy is located upstream of the one mentioned previously, but it too works toward the reduction of morbidity, albeit indirectly. The direct effect sought is a strengthening of the health system and of its resources : number and competency of personnel, appropriate material means (buildings, equipment, vehicles, etc.), available financing, community involvement in its own health development. The activities to be implemented here involve training, supervision, facilitation, construction, etc. (cf. figure 2).

Economic development aims at improving the standard of living of the population (this is the general objective), so it may devote more of its resources to its health. This strategy, much more far-reaching than the previous two, touches on a variety of sectors and should not be neglected : the opening of a road enabling a village to sell its production may have a greater impact on health than the opening of a clinic. Improvement of the standard of living depends on activities in a variety of productive sectors : farming, foresting, animal husbandry, fishing, crafts, trade and industry (cf. figure 3).

STEPS IN PLANNING

Before discussing the different types of evaluation, the steps in planning and action must be analysed. The planning process involves a number of steps :

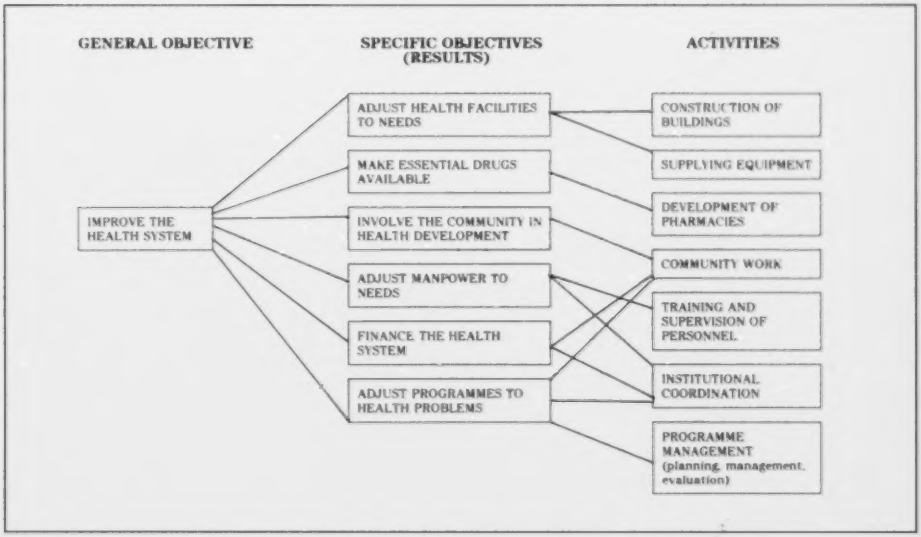


Figure 2 : Activities and possible results for health system support

This figure is read from right to left for evaluation (what result may be expected from these activities?) and from left to right for planning (what activities should be launched to achieve what objective?).

BEST COPY AVAILABLE

- identification of health problems and choice of priorities at which the programme will be directed (such as diarrhoeal diseases in children) ;

- choice of a strategy depending on the general objective : for instance, the reduction of mortality and morbidity caused by diarrhoeal diseases in children from birth to age 4 years ;

- the definition of specific objectives for the programme is the precise quantification of what results are expected from action : for example, a 75 % reduction in mortality caused by diarrhoeal diseases in children under 5, within 5 years, or a 25 % reduction of the incidence (new cases) of diarrhoeal diseases in children under 5 within 5 years ;

- a description of the activities to be set up for the attainment of these objectives. Each unit should direct its work at a target population and be capable of quantifying the services to be provided and the target population involved ; these may include educating mothers of small children and oral rehydration of diarrhoeic children at the consultation. The objectives may be more precise, such as : 50 % of mothers of children under 5 should be familiar with the preparation of the salt-sugar solution (oral rehydration solution), or the rehydration of all diarrhoeic children seen in consultation in accordance with their degree of dehydration ;

- quantification of the means to be implemented for the functioning of activities : human resources (personnel and community), financial resources, material resources (posters, oral rehydration solution, solution for infusions, vehicle for field visits, etc.).

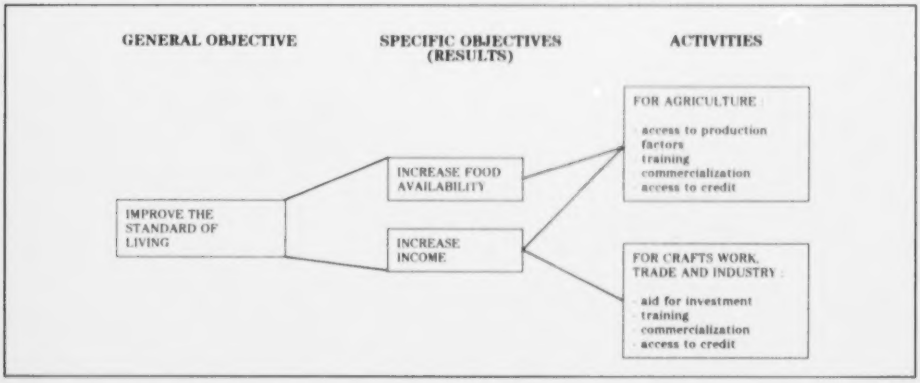


Figure 3 : Activities and possible results for economic development

This figure is read from right to left for evaluation (what result may be expected from these activities?) and from left to right for planning (what activities should be launched to achieve what objective?).

Action follows the same pattern as the definition of the programme (cf. figure 4), but in the opposite direction. The resources actually available are those allotted to the programme for its functioning. They make it possible to run activities and offer services which will cover all or some of the target population. These activities will produce effects or results pertaining only to the specific objectives (direct effects), which may be totally or partially achieved. Other effects will occur as well, some expected others unexpected, and the sum of these is known as impact.

To evaluate means first to ask questions, then to attempt to answer them. Evaluation's original question about the value of action or the expected results will be gradually refined at each step in planning and action.

Before action is begun, there may be some doubt as to whether the programme that has just been designed is coherent and realistic : will its objectives lead to the solution of high priority problems? Do the planned activities have a chance of achieving the objectives? Are the allotted resources sufficient for all of the activities?

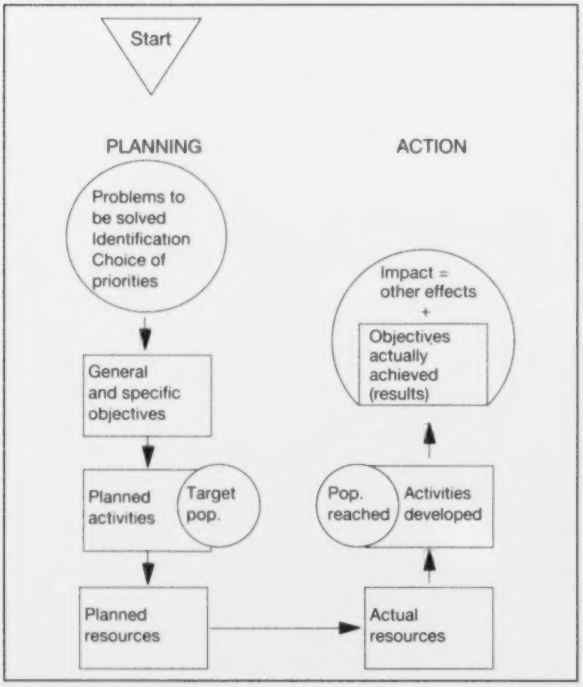


Figure 4 : Steps in planning and action

This evaluation of planning is an integral part of planning itself.

Evaluation of resources

Evaluation of activities

Evaluation of implementation

During an intervention, three series of questions may arise. The first pertains to the evaluation of resources : do the means actually engaged in the action correspond to what was planned? Are the promised resources available? The second involves the evaluation of activities or of the "process" : do the activities actually set up correspond to what was planned? The third has to do with the evaluation of implementation : are the target population groups actually reached by the programme? The term evaluation of implementation, or again, evaluation of deviations - may also be employed for these three series of evaluations.

If the action was launched without strict prior planning, evaluation remains possible on the basis of implicit norms : at each phase, an estimation is made afterward on the basis of what seems to be a normal or correct achievement.

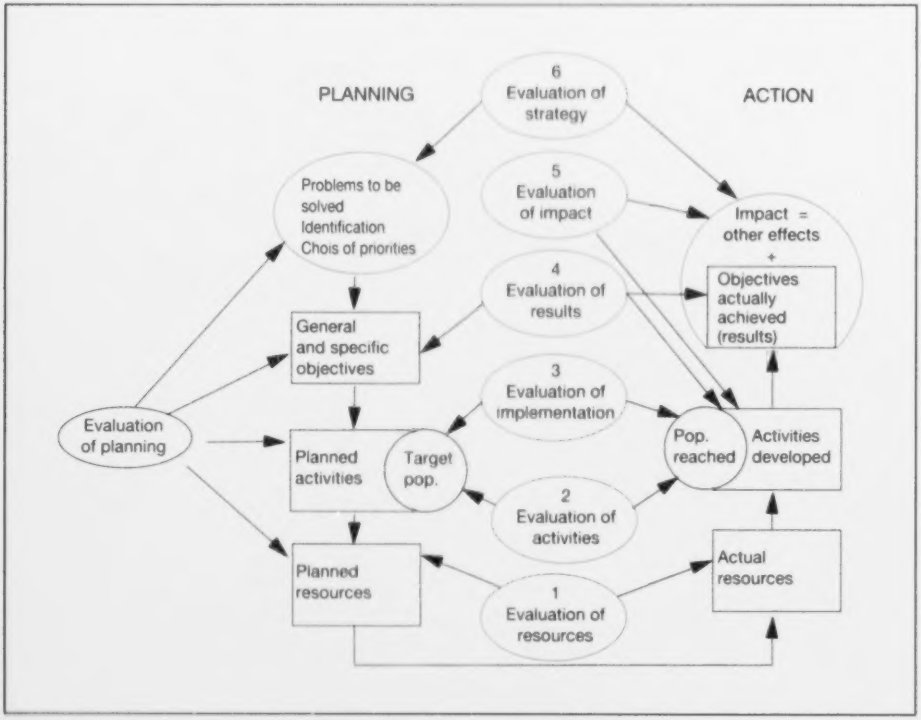


Figure 5 : Phases of evaluation

These three types of evaluation, which are most important for training purposes, show what portions of the programme have really been applied. Steering of a programme requires that they be performed at short intervals (quarterly or annually).

Evaluation of results

Evaluation of impact

Evaluation of strategy

When the action is finished, or following a relatively long period of time, the effects of the programme are investigated. Evaluation of results analyses the effects attained by activities implemented, and determines whether the specific objectives have been achieved. It is preferable that analysis be extended to the broader consequences of the programme for the population involved. This is evaluation of impact.

There remains one last question pertaining to the impact of the programme and its adequacy to the health problems requiring solution. This is evaluation of the strategy, or analysis of the real (or empirical) relevance, since it is based on the actual experience of action in the field.

For summative purposes, evaluations of the programme's effects are the richest of all, but they are meaningless until some time after the action, when the main points of the programme have been developed and the target population reached. The periodicity of these evaluations is somewhere between one and five years.

Evaluation of planning

Evaluation of action

Evaluation of effects

All of these evaluations (cf. figure 5) may seem discouragingly complicated, but they actually fall under three headings. Evaluation of planning means making sure that the different elements of the programme are coherent before it is implemented. Then, evaluation of action involves checking on the difference between what was planned and what is done, with respect to resources (evaluation of resources), activities (evaluation of activities) and target population groups (evaluation of implementation). Last, evaluation of effects studies the changes induced by the programme by comparing the expected results with those achieved (evaluation of results), investigating the overall impact of the programme (evaluation of impact) and reflecting on the strategy used (evaluation of the strategy).

Self-evaluation exercise n° 2.

1. Define three different community health action strategies.
2. What is the difference between an objective and an activity?
3. List three "action"-centred types of evaluation.
4. List three types of evaluation centred around the effects of the programme.

Answers at the end of this issue.

EVALUATING ACTION

Three types of evaluation covering resources, activities and implementation are dealt with here.

Each type of evaluation corresponds to a specific question aimed at determining whether the means invested were consistent with what was planned, whether the activities were set up according to plans, and last, whether the target populations were actually reached by the programme.

These evaluations are essential for conducting programmes, and must be done every three to six months, or at the most, once a year.

EVALUATION OF RESOURCES

The resources considered are both financial, material and human. Where finances are concerned, analysis includes funds coming from the government and aid from outside agencies, but also from the community. Material equipment (building facilities, furnishings, medical equipment, drugs, vaccines, facilitating material, vehicles, fuel, etc.) is divided into durables, that is, goods that are usable for several years (buildings, vehicles, furnishings, instruments, etc.) and perishables, which are used only once or for a short period of time (syringes, needles, compresses, reagents, drugs, notebooks, pens, fuel, etc.).

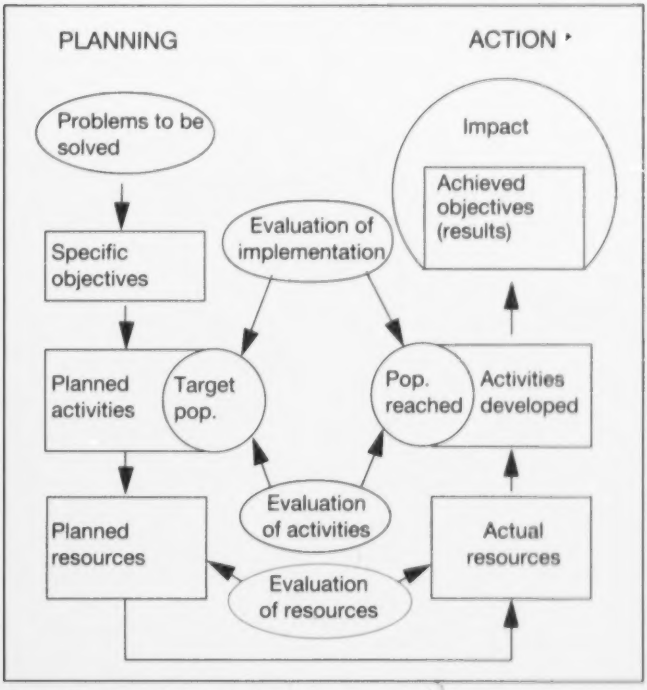


Figure 6 : Action-centred evaluation

As for the human aspect, the time worked by all protagonists in the programme must be considered irrespective of their level of training (community health agents, nurses, nursing aides, laboratory workers, doctors, midwives, maintenance workers, etc.) and their status (salaried, volunteers, semi-volunteers). The beneficiaries of the health action should also be considered, since their participation in activities or in use of health services also involves an expenditure of time and possibly of money.

Evaluation of resources consists of measuring the gap between those

resources actually available and those planned by the programme. This is easy to do, since the book-keeping is kept up to date, the equipment is inventoried (ordering forms, stock forms) and the personnel is known.

In those activities for which the resources required are proportionate to the work done, evaluation of the structure yields a first approximation. For instance, a vaccination centre used 100 doses of BCG, whereas the planned number for that period was 110. It seems probable that the centre functioned well. Caution should be exerted, however, since there is no proof that those 100 doses actually correspond to 100 vaccinations. There are always losses, vaccines left over at the end of a session and discarded, broken vials and refrigerator breakdowns. Conversely, the use of no more than 10 doses of BCG does not necessarily mean that a health centre functioned poorly, since the cause may be a shortage of vaccines at the supply centre or a problem in ordering them.

There is no automatic link, then, between use of resources and furnishing of the corresponding service. Evaluation of resources is no doubt the simplest type of evaluation, but it is also the most imperfect. Nonetheless, it may still be expected to yield justification of the use of resources ; that is, a rapid overall assessment of activities consuming material goods, such as vaccination, pharmacy and construction. This is a first approach to the causes of the gap between plans and actual use, since insufficient resources probably lead to a disturbance of activities.

How can information on available resources be obtained? A field visit will show the building facilities, the durable goods (an inventory of which should be made when necessary), the perishables (using invoices, ordering forms, stocks, stock forms) and last, the sums of money (book-keeping, cash box).

It is easy to measure use of personnel if this is confined to counting individuals, but it is preferable to establish a ratio between the number of individuals in each category and the population at large, representing the number of inhabitants per nurse or community health agent. It is more difficult to quantify the time spent at certain duties : this requires that schedules of actual activities be filled out, or an investigation made.

Available resources were estimated during planning. If not, a reconstruction of them should be made in accordance with planned activities and the target population.

EVALUATION OF ACTIVITIES (OR OF THE PROCESS)

Process evaluation pertains to the activities set up in the framework of the programme, and in community health they are usually based in MCH consultations, medical consultations, education for health, personnel training or community work.

Process evaluation involves measuring the gap between planned activities and those accomplished, in quantitative and qualitative terms. Planned activities may be specified by defining what action is to be done, at what rhythm, and possibly the level of quality. For

instance, conducting an immunization session one morning a month in a village or weighing children under three during two mornings a week (quantity) and referring the undernourished to a nutrition demonstration (quality).

Evaluation of activities provides direct measurement of programme "output", whereas evaluation of resources only provides indirect indications. This may show that the mobile vaccination team conducted 12 of the 15 sessions planned for the month, while three were cancelled because of rain, or else that two of the eight monthly nutrition demonstration sessions actually were held, and the other six cancelled because mothers did not attend them. Quantitative evaluation of the process is easy, since the object to be "measured", a meeting or consultation, is quite real.

Evaluation of quality is more difficult but contributes important information for the assessment of how the activity performed : what use is it to weigh infants and establish their growth curve if these are not interpreted, and no action is undertaken when curves stagnate or even tend downward?

Evaluation of the quality of activities is based on criteria defining those acts which should "normally" be performed (or possibly, those which are proscribed). A list of the criteria is made : it should be relatively short, so that each point may be examined, and sufficiently complete, so that all important aspects may be reviewed.

For immunization, criteria for quality involve making sure that vaccines are stored in a refrigerator at a temperature of 0 to 8 °C, that they are really available at each session, determining whether they conform to the schedule, whether a sterile needle and syringe (boiled for at least 20 minutes) are used for each child, whether the record card is correctly filled out, with the date and batch number, whether the date of the next vaccination is communicated to the mother and whether activities reports are sent out every month. This list may be completed by the enumeration of exceptions resulting in the non-respect of criteria. For instance, given the criterion that the vaccine be available at each session, the exception may be a shortage at the supply centre.

Evaluation of activities requires that some information be collected. For quantitative analysis, the number of activities actually conducted must be determined : number of opening days for consultations, of vaccination sessions and of supervisory visits. All of this information is available from the people in charge of the activities, who keep on-going records of them. These figures are compared with those formulated in the plan or programme agenda.

As for the quality of activities, the list of criteria is established on the basis of specialized norms. When these are lacking, the list is built by enlisting competent individuals in the sphere to be evaluated.

Actual field-work data may be obtained by a number of methods. The equipment may be observed, or certain measurements made. For instance, the cold chain may be evaluated by using cards equipped with a temperature indicator, or by checking the cleanliness and orderliness of the refrigerator. Some questions about transportation of vaccines or the immunization schedule may be put to the people in charge. Observation of the personnel at work, or for that means, of the activities reports, may yield some relevant information. Last, it may be useful to extend the investigation to users : this is a more cumbersome process, but it yields valuable insights as to the care-giver/receiver relationship and the perception of services offered.

However, process evaluation is of limited value : how important is a high-quality vaccination session if only 10 of the expected 50 children attended? What conclusion may be drawn from a training period attended by 5 individuals instead of the 15 invited? This evaluation should therefore always be completed by an evaluation of implementation, to determine whether the target population has really been reached.

EVALUATION OF IMPLEMENTATION

The target population is not the same for different activities : general consultations are for the entire population of an area, while antenatal consultations are for expectant mothers and garden plot-weeding is for families. The question of whether the target population has been reached is therefore necessarily addressed for each specific activity. It may disclose the fact that 30 pregnant women out of 45 are registered with the antenatal consultation, that 12 nurses out of 19 received training and that 80 out of 120 concessions have toilets.

Measurement of the population reached by a programme is easy if there is a direct interpersonal relationship, as in a consultation, a training course or a supervisory visit. In this case all that is needed is a record specifying the criteria defining the target population, such as dwelling place, age, sex and occupation. Users of the antenatal consultation are pregnant women usually residing in the area : those who are residing there temporarily are not counted, nor are those who believed themselves pregnant but were not. Sometimes the population reached cannot be found on records ; this is the case during an education for health campaign using posters or radio broadcasts, since there is no way of knowing who received the message. In some mass immunization campaigns no individual record is used and there is no knowledge of which vaccines were given to each child. A population survey must then be made. In the former case, a UAB (Understanding-Attitudes-Behaviour) survey is required, whereas the second should investigate vaccinal coverage.

Here too, measurements of quantity may be completed by an assesment of quality. If a single antenatal consultation or annual infant weighing is deemed insufficient, criteria for quality such as at least two antenatal consultations or at least one quarterly weighing are defined. Attention is paid here to the quality of the parti-

cipation of the target population, which is complementary to the evaluation of the quality of activities. Expected implementation should be defined when activities are planned, and criteria for quality should be specified when necessary (for example, 50 % of children aged 12-24 months will be vaccinated against measles and 70 % of mothers of undernourished children detected at weighing will participate in at least two nutrition demonstrations.

The target population is easy to quantify when it is small, as in the case of health workers or nurses requiring refresher courses. It is more difficult, however, when the entire population, an age group (children under 5, for instance) or a specific group (pregnant women) is involved. The ideal is to have a recent, reliable population census that is broken down for villages and hamlets and age groups, including children under age 1 and aged 1-4 years (1). If it is less than five years old it may be up-dated with the use of the natural population growth rate (NPGR) :

Population at $t + n$ = (population at t) $(1 + NPGR)^n$

Example : population at t (time of census) = 3,000 NPGR = 3 %

Population at $t + 4$ (4 years after the census) = $3,000 (1 + 0.03)^4$
= 3,000 x 1.1255 = 3,377 inhabitants

If the country's natural population growth rate is not known, it may be approximated using figures of 3 % for Africa and 2 % for Latin America and southern and southeastern Asia.

Beyond 5 or 10 years, extrapolation of census figures becomes questionable because of migratory movements. Inferences must then be made by cross-checking with other sources. Administrative counts for tax purposes are generally underestimated, while those done for electoral purposes or for implanting infrastructures may, conversely, be overestimated. In sparsely populated rural areas, local leaders and health workers are often personally familiar with the entire population. There also remains the solution of conducting a survey, possibly coupled with another one done for health reasons. The best solution is to count households (concessions, plots), then to take a sample of these, in which the number of inhabitants (5 to 7 inhabitants per plot in Africa) is counted. The average figure, multiplied by the number of concessions, shows the overall population.

The above-mentioned methods are also valid in the case of age groups. If no local age pyramid is available an approximation may be based on the age distribution shown in table 1.

---

(1) 1-4 years = from first birthday to 4 full years, that is, the day before the 5th birthday.

Table 1

Standard age distribution

Ages	Africa	Latin America	Southern and southeastern Asia
< 1 year	4 %	3 %	3 %
1- 4 years	13 %	10 %	12 %
5-14 years	27 %	24 %	26 %
15-44 years	42 %	45 %	44 %
45-64 years	11 %	13 %	12 %
> 65 years	3 %	5 %	4 %
TOTAL POPULATION	100 %	100 %	100 %

Source : according to United Nations figures, 1985.

An example in Africa : population = 3,000 inhabitants

Number of children under one = 3,000 x 4 % = 120

Number of women aged 15-44 = 3,000 x 42 % x 1/2 = 630
(the number of men and women is assumed to be equal).

For children under one, an estimation of number of births based on the vital events registers or the birth rate may be used. Vital events register figures are apt to be underestimated because of deliveries at home or outside of the area. They may also be overestimated if women come from other areas to give birth there. Country-by-country crude birth weights are published by the United Nations and the World Bank (1).

To convert the number of births into the number of children under one, infant mortality must be taken into account (mortality in children under one year of age). However, the figure used is only one-half of the infant mortality rate (IMR), since the basis for calculation is the number of surviving children at the middle of the year :

$$\text{No. of children} < 1 \text{ yr} = \text{annual no. of births} \times \left(1 - \frac{IMR}{2}\right)$$

If no recent national or regional figures are available, an approximation may be reached using the figures in table 2.

(1) Annual number of births = total population x crude birth rate.

Table 2
Average birth and infant mortality rates

Rate	Africa (‰)	Latin America (‰)	Southern and southeastern Asia (‰)
Crude birth rate	46	32	34
Infant mortality rate	110	60	100

Source : according to United Nations figures, 1985.

An example in Africa : population = 3,000 inhabitants

Crude birth rate = 46 per thousand

$$\text{Annual number of births} = 3{,}000 \times \frac{46}{1{,}000} = 138$$

Infant mortality rate = 110 per thousand

$$\text{No. of children} < 1 = 138 \times \left(1 - \frac{110}{2 \times 1{,}000}\right) = 138 \times 0.945 = 130$$

To estimate the number of pregnant women, the annual number of births is used, assuming that twin births approximately cancel out foetal deaths : the annual number of pregnancies is the same as the annual number of births. This figure therefore represents the number of women whose pregnancy ended during the year. To determine the number of "visibly pregnant" women at any given time, however, the annual figure must be divided by two (since women are "visibly pregnant" for about 6 months, from the 4th to the 9th month).

To sum up : there are a number of sources available for the evaluation of implementation. Measurement of the target population reached by the programme requires a clear definition of that target population, as well as records of use of health units, or else a survey of the target population.

Estimation of target population groups requires a recent population census, knowledge of the natural population growth rate, the age pyramid, the crude birth rate and infant mortality rate, some administrative data, information supplied by community leaders and health personnel and possibly by a population survey.

CHOICE OF INDICATORS

Direct health action

In the direct health action strategy, evaluation of action means evaluation of the health services set up in the programme framework. These evaluations are very important in determining whether the population is actually covered by the health units.

In the case of education for health, evaluation of action involves measurement of the diffusion of messages : number of sessions held, of people attending, of posters put up, of radio or television broadcasts. Evaluation of quality looks into the educational methods used, the form of the messages (attractive material, clear message) and the quality of people's participation (questions, remarks).

Evaluation of implementation makes sure that the message was actually received. A short survey of a few participants at the end of the session yields some information. When working with the mass media, a home survey of knowledge conveyed by the messages broadcast is needed. This appraisal is quite superficial : for instance, should children be vaccinated against measles?

Analysis of attitudes is a more difficult phase. It deals with the way in which the messages interact with the value system of the individuals reached. This value system lends positive or negative colouring to messages, but attitudes are not set once and for all, and may change. For instance : some people say that Aids is a sort of punishment for wrong-doing, do you agree with that interpretation?

This survey for the evaluation of implementation is generally combined with one on the results of education for health, concerned with behaviour modification, to form a UAB (Understanding-Attitudes-Behaviour) survey.

Health development

In the health development strategy, there are specific indicators for each activity. Training and supervision are often the main actions. For actions aimed at building and equiping facilities, evaluation concentrates on whether the schedule and budget have been respected, as well as on the quality of the work done. Far charts are a good way of visualizing the different stages : each one is represented by the outline of a bar, the length of which indicates the planned duration. The outline is filled in as work is accomplished, for evaluation of how building advances, for instance (cf. figure 7).

Be they private, public or community-run, pharmacies theoretically keep a list of essential drugs, which must be distributed and stored. Control and management make sure of proper use of financial resources for restocking the pharmacy and covering running costs.

Criteria for evaluation pertaining to essential drugs involve the number of products (no more than 20 for a health outpost and 50 for a health centre), their relevance to local pathology and the competency of the personnel. Drugs are judged for their effectiveness, absence of toxicity, ease of use and low cost.

For distribution and storage, the pharmacy's monthly turnover is the best indicator. For qualitative analysis criteria pertaining to distribution may be used : respect of opening hours, correct handling of tablets, accurate reading of prescriptions, accurate noting of

income. As for storage, points such as cleanliness of the storage room and whether it is locked, the existence of a storage cabinet that may be locked, marking of boxes that will expire within the year, sale of drugs in accordance with their expiry date and updating of stock records may be considered.

Control and management positions will be evaluated by the number of checks on the cash box and reserves done or planned (the cash box should be checked at least twice a month and reserves once a month), the number of accurate controls and the number of controls performed (evaluation of the quality of distribution and storage), the number of orders made and planned (generally one each quarter). It may also be useful to make sure that there is enough money available to cover running costs and restock the pharmacy.

For the community, facilitation fosters a self-teaching process of identifying problems and seeking solutions. This is a delicate, innovative activity, and is therefore rarely brought to its conclusion. It requires local administrative autonomy, a modicum of health services and group leaders accepted by the community. There are several methods such as participatory surveys, or group sessions using flanellgraphs. They are based on meetings of the entire community or of homogeneous subgroups such as women or young peasants. The creation of a formal community development structure such as a committee for the promotion of health is not compulsory.

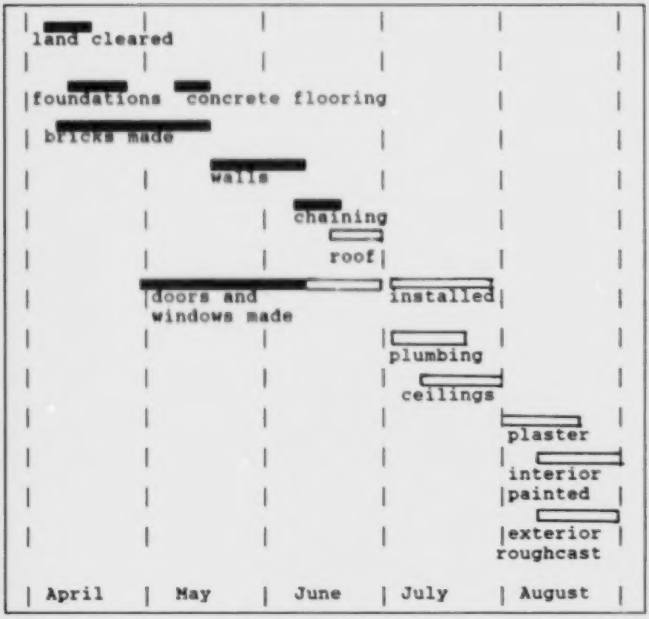


Figure 7 : Evaluation of progress in construction, using a bar graph. (the bars filled in correspond to the situation on 30 June ; the work in therefore two weeks behind schedule).

Criteria for evaluation may include the number of actual and planned meetings, the number of actual and expected participants, the positive attitude of local authorities and health personnel, what was said during the meetings, the extent of agreement in analysing the situation, decisions made at the end of the meeting and the calendar for implementing them.

DECISIONS MADE FOLLOWING EVALUATION

Evaluations centred around the action itself yield an accurate picture of the degree of advancement of the programme with respect to use of resources, activities initiated and implementation with respect to the target population groups. The value judgement is made by comparing these results with what was originally planned, keeping in mind the fact that the plan is not something to be achieved at all costs, but rather a scale against which to measure achievements, and one which may in fact turn out to be erroneous, through either excessive optimism or pessimism. If the planned activities have not been accomplished, the reasons should be sought, for the benefit of the next planning-action-evaluation cycle.

In formative evaluation, evaluation of action may result in different decisions such as the adjustment of resources or of activities.

In summative evaluation, evaluation of action may lead to the measurement of effects if the programme has been implemented and has touched the target population, or on the contrary, it may cause the evaluation of effects to be delayed, or even call the programme to a halt.

LIMITS OF ACTION-CENTRED EVALUATION

In every case, the central question in evaluation is to determine whether action has really improved community health. When evaluating action, there is a desire to push conclusions further. For instance, in one programme 100 of the planned 130 doses of measles vaccine were used (evaluation of resources) during the 7 actual sessions, for 8 planned (evaluation of activity). 84 children were vaccinated, 77 of whom belonged to the target population of 9-12 month-olds, estimated at 100 (evaluation of implementation). The quality of the cold chain and the respect of the immunization schedule seemed to be satisfactory (evaluation of activity). Does this mean that vaccinal coverage for measles is good and that the immunization programme is effective? In the present case, the answer is "definitely, yes", since experience with immunization has shown that properly administered vaccination of children really affords close to 100 % immunity.

This is not true for all types of action, however. In a malaria prevention programme distributing chloroquine to pregnant women, ten education for health sessions were conducted (of the 12 planned), with the participation of 80 % of the pregnant women in the area. Does this mean that the women are protected against malaria? Even if more complete process evaluation is made, and if we find that the message was actually received by 70 % of the women, there is no indication that they will really take the chloroquine and thus be protected.

We have already seen that consumption of resources does not automatically correspond to implementation of an activity. Now we see that the implementation of an activity does not necessarily produce results. This shows the limits of action-centred evaluation : no matter how good it is, it rarely yields conclusions as to the effects of action. This may seem a truism, but the best way of

determining the effects of a programme is to measure these effects themselves.

Self-evaluation exercise n° 3

Alphonse is the nurse in charge of the leprosy circuit in the district of Obama, which counts 12,000 inhabitants in Obamo City and 6,000 inhabitants in four villages, A, B, C and D. 20 / of the population is believed to suffer from leprosy.

Here is Alphonse's programme :

- visit the four villages A, B, C and D and neighbourhoods I, II, III and IV of Obamo City once a month ;

- hold a public meeting in each place to explain what leprosy is and encourage people to go for screening. The meeting is only held if more than ten people are present ;

- at the end of the meeting, do initial screening on people who have noticed signs on themselves ;

- refer detected cases to see the doctor (new cases) ;

- last, at each monthly visit, see the cases authentified by the doctor (old cases) to give them an injection and give them tablets for the coming month. Treatment lasts 2 years.

Here is a record of Alphonse's activities for the first semester.

Villages or neighbourhoods	A	B	C	D	I	II	III	IV
No. of visits	3	2	4	3	6	5	5	6
No. of meetings with more than 10 people	3	1	4	2	6	4	4	5

Situation for the entire district	JAN.	FEB.	MAR.	APR.	MAY	JUNE
New cases detected by Alphonse	21	7	15	8	10	12
New cases then seen by doctor	4	9	11	8	4	5
Old cases on record in village or neighbourhood	62	58	65	74	76	79
Old cases seen and treated by Alphonse	58	45	63	56	73	70

Using these data, evaluate leprosy control action in the Obamo district on the basis of implicit norms.

Answers at the end of this issue.

EVALUATING EFFECTS

This heading includes three types of evaluation : of results, impact and strategy.

The idea is to determine the actual effects of action. These are the most important types of evaluation, since they are instrumental in making an overall assessment of a programme, but they are heavier and more costly than action-centred evaluations. They are only meaningful if the planned activities have been implemented to a large extent, barring which they should be delayed. It takes some time to set up a community health programme, and after that for the people involved to assimilate the messages and use the services offered. These evaluations of effects should be conducted after a period of one to five years (cf. figure 8).

EVALUATION OF RESULTS

Evaluation of results should show which specific objectives have really been achieved through the activities implemented. It deals with the direct effects expected of action on the target population. To take the action's coverage of the target population into account, the analysis may distinguish between utility and effectiveness of the programme ;

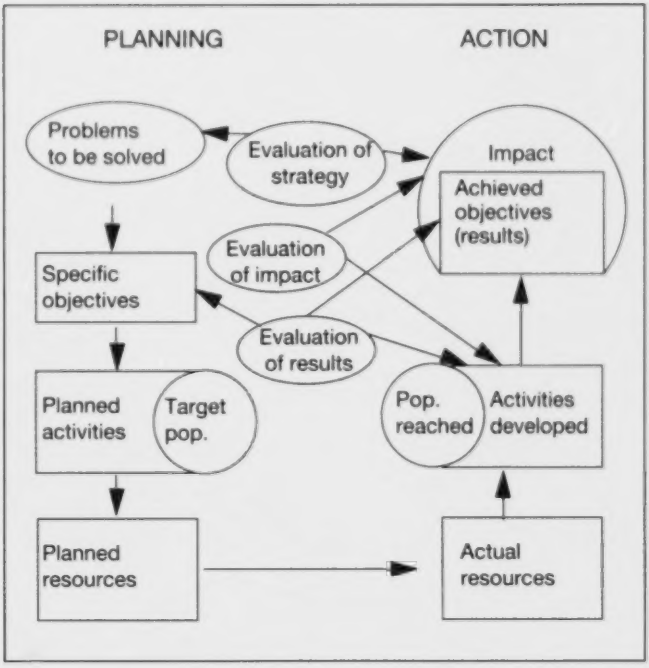


Figure 8 : Effect-centred evaluation.

- utility pertains to the effects of the programme on users. In other words, it measures the results that would have been attained if the entire target population had been covered : it is therefore a measurement of the theoretical effectiveness of the actions within the programme. It is valuable for actions for which the results cannot be foreseen, such as education for health or training. Its value is slight when the utility of the actions is well-proved and constant, such as vaccination ;

- effectiveness speaks to the effects of the programme for the entire target population irrespective of whether it was reached by the programme. This then measures the actual effectiveness, given the real acceptance and use of the services

by the population : an action may seem quite useful on paper, but will not be effective in the field if it does not reach the population at which it is aimed. Measurement of effectiveness is therefore the most complete evaluation of results.

Direct health action

In the direct health action strategy, the result of the programme is a change within the population ; this may be a change in behaviour, improved immunization or reduced mortality. Evaluation of results compares measurements made within the population with reference values.

Measurement within the population means making a survey. Once again, measurement confined to users of services evaluates utility. To evaluate effectiveness, measurements must be made among users and non-users. When the population is large, it is often more economical to do a sample survey. The sample must then be exactly representative of the population, and requires randomization or the use of pseudo-randomizing methods such as the itineraries method. The protocol known as "30 x 7" based on the random selection of 30 groups of 7 units, originally conceived for the evaluation of vaccinal coverage, has repeatedly proved its validity. It may be used for all sorts of surveys requiring an accuracy that does not exceed plus or minus 10 % of the estimated percentage.

Without going into the detail of health surveys, here are some indispensable prerequisites for high quality :

- the objective of the survey is specified (what variables are to be measured, what is the utility of determining them?) ;

- the population and the groups covered are well defined (for example, families, or mothers of children under 5) ;

- the size of the sample is adjusted to the degree of accuracy desired and to the available means ;

- the survey protocol is put in writing in concise form : choice of groups to be investigated, questions to ask, "measurements" to be made (weight, height, etc.), what to do in case of non-response (person absent or refusal to answer). When drawing up the protocol and exploiting findings it is preferable to obtain the help of a person trained in conducting surveys ;

- no more than 20 to 30 questions are asked in the usual tongue of those questioned, and each interview lasts no more than 30 to 45 minutes ;

- the authorities have given their agreement and informed the population ;

- the survey workers have met for prior training, and the first questionnaires were filled out in the presence of a supervisor ;

- the survey has been conducted before being launched.

Comparison with reference values raises the question of the choice of a control group. It is extremely difficult to resort to the experi-

mental method (randomization) to evaluate the effectiveness of a community health programme. Two other "quasi-experimental" methods with satisfactory validity may be used. First, before/after comparison, in which measurements at the end of the action are compared with those made beforehand. One good point here is that the population assayed is the same, but the shortcoming is the existence of factors other than the programme, which interfere and participate in any changes found. Measurement before, during and after the intervention is more valid, but also more costly.

There is also the here/elsewhere comparison, in which the measurements made on the population having benefitted from the programme are compared with those of another group that was not included in the programme, or from a neighbouring area. This method does not require assay before the intervention, but two population groups are rarely totally comparable. The choice of the reference area must necessarily be based on criteria independent of the programme : for instance, an outlying community that refused to participate in the programme should not be used as a reference.

Validity may be improved by combining the before/after and the here/elsewhere methods, but the expense is then greater. All other methods for selecting a control area involve observation only, and do not authorize the conclusion that the programme is really the cause of the change found in the population.

Health development

In the health development strategy, the direct effects expected pertain to the health system, including its facilities, equipment, drugs, financing, health manpower, community involvement. Evaluation of results is simpler here, since no survey is required. While the ultimate goal is still the improvement of the population's health, what is being evaluated is the impact (an indirect effect of the programme) rather than the results (direct effects of the programme).

Evaluation of results is concerned with the health system resources, beginning with the facilities, equipment and drugs. Are the buildings large enough, is the number of rooms sufficient for the work planned? Is the layout functional? Is the equipment adequate for the planned services? Are drugs and other perishables available all the time?

In the financial sphere, the rate of recovery of running costs is an indicator of whether the health structures are self-financing.

$$\text{Rate of recovery (\%)} = \frac{\text{total annual income}}{\text{total annual running expenses} + (\text{initial stock} - \text{final stock})}$$

Income may be generated by the sale of drugs, by services paid for by users, membership fees, "fines" or donations. Expenses include maintenance of equipment and the purchasing of perishables (drugs, syringes, dressings, reagents, fuel, paper goods,

etc.). The purchasing of durables is not considered, nor are the wages of workers who are civil servants.

A rate of 100 % means that continued provision of services is feasible. A lower rate is not viable. A higher rate leaves extra money for buying other equipment, increasing stocks or contributing to other development-oriented interventions.

Health manpower is analysed to determine whether the staff is large enough for the work demanded of it, whether it is competent and motivated for the accomplishment of its duties. This is done by evaluation of tasks.

To evaluate results, indicators of coverage by the health unit (availability, accessibility, utilization, adequate utilization, effective coverage) may be used. It is a fact that a unit provides proper coverage of the population if the equipment and drugs are available, if it disposes of financing and if the personnel is sufficiently numerous and does high-quality work.

Community involvement in health development is measured by the degree of its participation in initiating, organizing, performing and evaluating action. The number of interventions in which the community participated and the extent of its participation may be evaluated. Qualitative assessment may be added : is participation spontaneous, induced or "requisitioned"? Was the community a motor force, a partner or simply a work force?

EVALUATION OF IMPACT

What is at issue here is the diagnosis of the overall effect produced by the activities implemented. Whatever the strategy, evaluation of impact must pay attention to all of the effects induced by the programme, and more particularly to changes in the health status of the population, trends in behaviour relating to health, disease and the health system, modifications in community life, in its capacity for self-organization and mobilization, economic tendancies reflected in production and consumption. It is also interesting to analyse how the programme is perceived by the community.

While evaluation of impact may make use of surveys (UAB or morbidity-mortality surveys), it is nonetheless essentially qualitative, and tends to rely on the semi-directive interview technique. This involves questioning a relatively small number of individuals (10 to 30) but thoroughly enough to determine their true opinion of the programme and its effects. These individuals are chosen within the target population and among influential members of the community (administrative and political officials, prominent citizens, teachers, heads of trade unions or clubs, etc.). If these interviews are to be relevant, the interviewer must be good at communication.

EVALUATION OF STRATEGY

The goal of this phase is to analyse whether the impact of the programme was coherent with the problems to be solved.

Evaluation of strategy is a qualitative and quantitative comparison between high-priority problems and actual impact, in the light of data yielded by evaluation. This is the beginning of a new planning-action-evaluation cycle.

Decision-making following evaluation

In formative evaluation, the evaluation of the effects of a programme may lead to decisions on the adjustment of activities and resources if the expected effect has not been achieved, the challenging of activities and resources if the impact is far from affecting high-priority problems, or again, criticism of the objectives if evaluation shows a change in the problems requiring priority.

Summative evaluation may disclose a need to terminate the programme if it did not yield results, if too many perverse effects are seen, or again, if there is little impact on high-priority problems.

Conversely, the decision may be to prolong or extend the programme if the results yielded by the activities corresponded to expectations, or to discontinue the programme if the objectives have been achieved.

Self-evaluation exercise n° 4

Three years ago, the Highland district health team launched a programme to combat weaning-related malnutrition in infants. To do so, it organized nutrition demonstrations that were held in the five district health centres at the end of weighings for infants, so that many mothers would participate. Furthermore, leading women were given training in health centres so that they could amplify certain messages and actions in their village. These women were first chosen in the villages furthest from the health centre. Since care for malnourished children was unsatisfactory, a nutritional rehabilitation centre was finally set up in conjunction with the paediatric ward in the district hospital.

At the end of three years, all of these activities were doing well. The district health team and the regional director of health wished to extend this programme to the neighbouring district and raised some questions as to the effects of these activities.

How can their utility, effectiveness and impact be evaluated? In a few sentences, outline a protocol for these three evaluations.

Answers at the end of this issue.

BEST COPY AVAILABLE

SUCCESSFUL EVALUATION

Like any other health action, evaluation must be planned, performed and evaluated.

PHASES IN EVALUATION

To be successful, evaluation must go through a number of sequential phases similar to those required for a health action programme.

In the first phase the subject of the evaluation, the required level, method, information and resources are defined, and a protocol and calendar are drawn up. The evaluation is then performed - that is, data is collected and analysed and conclusions drawn and sent out to decision-makers. Before drawing conclusions, the validity of the evaluation must be assessed (evaluation of the activity) and its consequences determined (evaluation of impact).

Definition of the subject of the evaluation means determining what decisions should be made as an outcome of the conclusions drawn by evaluation, and deciding whether evaluation will be formative or summative. Participants, as well as the recipients of the conclusions, must be identified : are they the community, an administration or another agency?

Given the different levels of evaluation and the numerous possible indicators, there is no way of evaluating all aspects of a programme ; a distinction is therefore made between evaluations centred on action and those centred on effects.

Evaluation may also be "molar" or "molecular", using the vocabulary of chemistry to indicate difference in scale : molecular evaluation exercises only a small part of the programme, assumed to be representative of the whole (this is a sort of sample survey), whereas molar evaluation examines all of its parts.

The tracer method used in molecular evaluation consists of testing several interventions by following their different aspects (resources, activities, results, impact) so as to achieve an evaluation of the programme as a whole.

To establish a protocol for evaluation, indicators must be chosen, along with the information implied by them and how it is to be collected (survey, record of activities).

To allocate resources and establish a calendar one must determine who can participate and over what period, and identify the equipment and means available. Following this, a preliminary agenda for evaluation may be established, data collected on the basis of the predefined protocol and analysis performed with the help of statistical tools.

Evaluating the evaluation process is the equivalent of studying the validity of evaluation. The diagram of the programme's objectives must be examined to make sure that the indicators selected are relevant. Data collection is then checked to make sure that it respected the predefined protocol. Last, biases which may have dis-

torted measurement or confounding factors explaining any differences found should be sought.

A report is written containing the conclusions of this evaluation, for distribution to decision-makers. Three qualities are required of it : first, it should be available on time ; that is, before decisions relating to the programme are made. Second, it should be accessible ; that is, appropriate in size. If too long it will not be read, and if too much of a résumé it can only show a caricature of the situation. Its style and vocabulary should be comprehensible by its potential readers. Last, it should be generally accepted ; that is, not be too polemic. Evaluation includes a value judgement, with its positive and negative aspects, none of which should be hidden. If some extremely negative facts are uncovered during evaluation, they should be aired before the report is published, so that the latter may present concrete suggestions for corrective action.

Last, it is useful to analyse the impact of evaluation by making sure that the report and its conclusions have been read, the recommendations understood, criticized and followed up, and if not, to determine the obstacles. Another significant point is finding out what consequences evaluation has had at the grassroots level.

Summary of the ten factors contributing to successful evaluation :

- Evaluation is planned at the inception of the programme.
- The protagonists in the programme are the protagonists of evaluation.
- The goal of evaluation is well defined and accepted by the protagonists in the programme.
- Evaluation leaders have confidence in the protagonists of the programme, are open-minded and capable of making reliable judgements balancing out the positive and negative sides, and able to apply the rules of evaluation in a flexible way.
- Indicators used for evaluation are valid, objective, available and their cost is reasonable.
- Evaluation is prepared by on-going collection of quantitative and qualitative information.
- The protocol for evaluation and the schedule are drawn up before evaluation is begun.
- The cost of evaluation (time, finances, resources) is acceptable in proportion with the cost of the programme itself (no more than 10 to 20 %).
- The conclusions of the evaluation are ready in time, accessible and acceptable.
- Evaluation itself is evaluated with respect to its validity and impact.

EVALUATING A HEALTH UNIT

Several indicators are useful in evaluating coverage by a health unit, availability, accessibility, utilization and adequate utilization, quality and actual coverage.

When evaluating a health unit, an attempt is made to determine the percentage of the population covered by this unit. In fact, the term "coverage" includes several different notions : the personnel and equipment available to perform the various jobs, the population actually attending the unit and the quality of the service delivered. Coverage must therefore first be broken down into several indicators : availability, accessibility, utilization, adequate utilization, actual coverage. Next, we will see how these indicators may be grouped on a diagram for interpretation purposes.

The reader is again reminded that these indicators constitute an evaluation of action within a direct health action strategy, but they evaluate effects within a health development strategy.

AVAILABILITY

Availability of equipment

Availability is calculated by the percentage of time during which the resources needed by the unit are present. For those services that are consumers of equipment, availability is evaluated by the length of the period during which no shortage occurred :

$$\text{Availability (\%)} = \frac{\text{no. of months (or weeks) without shortage}}{\text{12 months (or 52 weeks)}} \times 100$$

For instance, the availability of an immunization unit is evaluated by the availability of vaccines, while the availability of a curative care consultation is evaluated by the availability of the main essential drugs, etc.

Availability of manpower

Human resources are usually most important. The easiest way of evaluating the availability of personnel is to calculate the size of the population per health agent in each category. For instance, one doctor for 20,000 inhabitants, one nurse for 2,500 inhabitants or conversely, the number of health agents per 100,000 inhabitants (in this case, 5 doctors and 40 nurses for 100,000 inhabitants).

Unfortunately there are no international norms against which to compare these figures, since situations vary too much with the country's economic status, population density, urban spread, demands for care and the productivity of the personnel. If no reference is defined by the ministry of health at the national level, a reasonable norm for the region may be defined on the basis of available information on similar countries and on the situation of the most privileged quarter of the country's population living under similar conditions, be they urban or rural.

Observation of the demand for care is an easy way of evaluating availability : it simply requires calculation of the length of waits at general medicine consultations or for monitoring of children's

growth. However, it is not accurate, since availability will be just as good in a unit with a small staff, offering few consultations, as in one in great demand with a large staff. This method is only valid for measurement of the rate of use of maternity and hospital beds.

A third method consists of calculating the number of FTE (Full Time Equivalents) required to provide services, on the basis of the objectives set for services to be offered.

$$\text{Availability (\%)} = \frac{\text{number of FTE presently worked}}{\text{number of FTE needed}} \times 100$$

To estimate the number of FTE needed, the number of annual working hours required for the accomplishment of the work must be calculated. This is done by determining the size of the target population, the desirable number of contacts (4 antenatal visits and one delivery per pregnant woman, 5 immunization contacts per child under one, etc.) and the time (in hours) required for each type of act - that is, the productivity of the personnel. Now productivity may be extremely variable : during an epidemic, a nurse may see up to 200 patients a day (2 min. per patient), whereas in normal times a figure of 30 to 40 (10 to 15 min. per patient) would be correct. Here too, the lack of norms is a problem, since productivity depends on working conditions and available equipment to a large extent. Questioning and observing workers are two ways of getting some notion of how much time is required for each act.

Number of working hours = target population x number of contacts x time spent per act.

Calculation of the number of FTE is based on the rule that a working day includes 6 hours devoted to care, with the rest used to maintain equipment, make reports and coordinate activities. There are 250 working days in a year, representing 1,500 working hours a year.

$$\text{No. of FTE needed} = \frac{\text{No. of working hours a year}}{1{,}500}$$

In a population of 3,000, the number of general medical consultations per inhabitant and per year is estimated at 3, lasting 15 minutes each.

$$\text{No. of FTE needed} = \frac{3{,}000 \times 3 \times 1/4}{1{,}500} = 1.5 \text{ FTE}$$

At the health centre, a nurse devotes 50 % of his/her time to general medical consultations and a traditional midwife 30 % : 0.5 + 0.3 = 0.8 FTE.

Availability = 0.8 : 1.5 = 0.53, or 53 %.

This method is accurate, since it takes into account the specific work to be done, but it assumes that the entire target population

is covered, which is not often the case. The number of acts may then have to be modulated on the basis of actual attendance of the unit.

Last, there is a fourth method taking actual morbidity within the community into account, but it requires too much information to be easily applied.

Units offering curative care and childbirth services must be open 24 hours a day, while those monitoring mothers and children should provide consultations at sufficiently frequent set times.

ACCESSIBILITY

People encounter a number of obstacles, which may be physical (distance, travelling time), financial (cost of the service, cost of the trip) or sociocultural, in attending a health unit. For instance, a gynaecological consultation run by a male nurse is not very acceptable for women in some countries.

Geographic accessibility, the most important factor, may be calculated on the basis of the population living in a given perimeter :

$$\text{Accessibility (\%)} = \frac{\text{pop. living in a given perimeter}}{\text{total population of the area}} \times 100$$

As a rule, a one-hour walk (or 5 km on flat ground) between one's home and the health centre is viewed as the maximum acceptable for accessibility. For a community health agent or a mobile unit, the perimeter of accessibility is about one-half hour. A hospital should be less than two hours away by a vehicle. However, the cost of transportation (economic accessibility) must also be taken into account here. It may be compared with the minimum wages for a day's work, or the population may be questioned. Geographic accessibility may be visualized by drawing a circle with a 5-km radius around each health centre on a map of the area.

UTILIZATION

Utilization corresponds to attendance of the units by the target population :

$$\text{Utilization (\%)} = \frac{\text{no. of people having used the unit}}{\text{target population}} \times 100$$

Of course only users living in the area are counted. This indicator is usually calculated on an annual basis. For instance, 300 childbirths were expected in the course of a year, 200 took place in the maternity. Utilization of the childbirth unit was therefore 200 : 300 = 0.67 or 67 %.

For antenatal consultations, the utilization rate is the number of women registered with the consultation compared with the expected number of pregnancies for the same period, while for immunization the number of injected doses of each vaccine is divided by the putative number of children under one for the same period.

For curative medical consultations, the target population is not accurately known. According to some estimations, there is an

annual average, per inhabitant, of 2 to 4 episodes of disease requiring recourse to the health system. In the absence of references, only the utilization rate is calculated ; that is, the annual number of patients divided by the number of inhabitants (cases/inhabitants/year). A unit is insufficiently used when this figure drops below 0.5 in urban areas and 0.2 in rural areas.

It may be interesting to check on the physical accessibility of units by calculating the utilization rate with respect to distance from users' home.

Utilization is the most polyvalent criterion for evaluating a unit. A unit is utilized if it is sufficiently available, accessible, accepted, and if its reputation is good. If only a single indicator can be used, utilization is chosen.

ADEQUATE UTILIZATION

Adequate utilization, or adequate coverage (AC) measures high-quality utilization by users, such as a minimum of 6 visits for infants during the first year of life, or full treatment of a disease.

$$\text{Adequate coverage (\%)} = \frac{\text{no. of people who made adequate use of the service}}{\text{target population}} \times 100$$

Quality, effective coverage

The quality of services afforded by a unit must be assessed from the outside, by a supervisor : a survey of users may be valuable. A rating of overall quality expressing the proportion of acts of acceptable quality is calculated :

$$\text{Quality rating (\%)} = \frac{\text{no. of high quality acts}}{\text{total no. of acts}} \times 100$$

Thanks to this rating, the effective coverage of the target population by the unit may be expressed :

Effective coverage = adequate coverage x quality rating

Effective coverage is the proportion of the target population that made adequate use of the unit and received high quality service. It is the best indicator for evaluating a unit.

The definition of criteria for quality may be a long process. In a curative medical consultation, the supervisor may observe how the register is kept, how questioning and examination are done, whether diagnosis is accurate, whether therapeutic protocols are respected and the nature of patient-worker relations. For children, whether routine weighing is done and vaccinations checked whenever a child is seen for a disease may be indicators of quality.

In an antenatal consultation, checks may cover questioning, weighing, measurement of uterine height, blood pressure, checking for oedemas, examination of the conjunctivas, abdominal palpation, listening to foetal heart sounds (from six months on), testing for albuminuria and glycosuria. It may be worthwhile to analyse the criteria for referral to another level of responsibility when the case is beyond the competency of the local personnel.

For childbirth, are the rules for asepsis respected, are the pulse, blood pressure, foetal heart sounds and dilatation monitored systematically? The percentage of early post partum infections and of perineal tears are also indicators of quality.

For vaccination, the cold chain must be maintained, injections must be sterile and the immunization schedule respected.

As for the surveillance of infants, the observer should make sure the scales and measuring rods are reliable, and examine the methods of measurement and the interpretation of data

THE COVERAGE DIAGRAM

All of the indicators analysed reflect the level of coverage of the population in different ways. Since they form a coherent whole, they may be grouped on a graduated diagram of percentage of the target population. The availability of the unit, its accessibility (generally multiplied by availability to account for the influence of availability on accessibility), its utilization and adequate utilization are recorded one after the other, without omitting effective coverage, which is adequate utilization multiplied by the quality rating.

A curve is drawn, starting from 100 % of the target population. Its slope tends more or less upward and to the left. Ideally, if the entire target population was actually covered by an available, accessible, well utilized unit, the curve would be vertical at 100 %.

The coverage diagram is useful in identifying the main disturbing factor, which is the one for which the curve is most horizontal. In the present case, it is utilization. If it is availability or accessibility, an increase in the health system's resources is required. If it is utilization or adequate coverage, there is a problem of acceptance of the programme by the population, and education for health should be reinforced. If it is effective coverage, the quality of care should be improved : the competency and motivation of personnel should be enhanced through training and supervision.

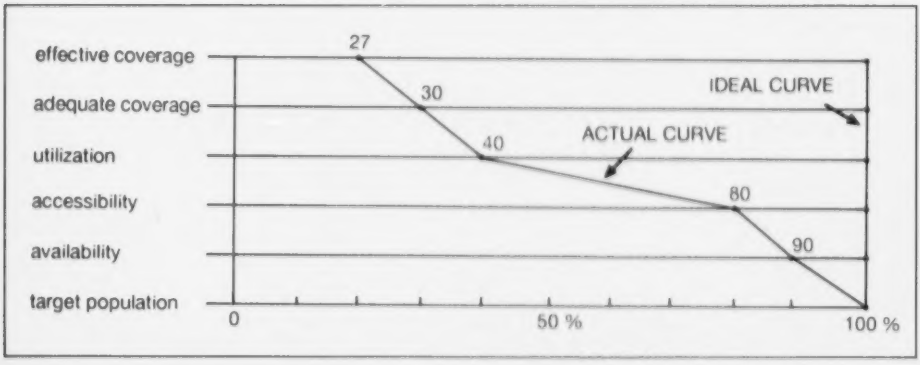


Figure 9 : Coverage diagram

Self-evaluation exercise n° 5

The district of Pescara has 5 health centres located in the five largest towns.

Location	Population	Population within a 5-km range
County-seat	20,000	80 %
Village A	5,000	70 %
Village B	8,000	50 %
Village C	4,000	50 %
Village D	1,000	90 %

In the past year, the district supervisor reorganized antenatal consultations (ANC), and the ANC is now open one day (6 hours) a week, with a tenured nurse in each village and two nurses in the county seat. During their holidays they are replaced, so that the ANC continue normally. Each pregnant woman should be seen four times during her pregnancy, but two visits constitute minimum coverage.

At the least, the antenatal consultation should include a talk with the woman, who is questioned, weighing and taking of blood pressure, measurement of uterine height and testing for albuminuria. The first ANC takes about one-half hour and fifteen minutes suffice for the following visits. The birth rate is 40 per thousand in the district of Pescara. The regional physician wishes to evaluate the ANC service in the district, and has asked the supervisor to make a report. The latter goes to the health centres and studies the ANC records of women who gave birth within the last six months. He counts the number of record cards (equal to the number of women who attended at least once) and the number of women who came twice or more. He then checks on whether the different pieces of information are accurately noted on the card and counts incomplete cards.

Six-month report	Total no. of ANC cards	No. of women who consulted twice or more	No. of incomplete cards
County-seat	323	247	45
Village A	72	54	18
Village B	101	48	19
Village C	48	23	8
Village D	31	21	1
Total	575	393	91

1. Do the evaluation requested by the regional physician (do overall calculations for the district as a whole).
2. Present findings on a coverage diagram.
3. What suggestions do you convey to the regional physician for improving the antenatal consultation unit?

Answers at the end of this issue.

EVALUATING TRAINING AND SUPERVISION OF PERSONNEL

The objective of training is the improvement of the personnel's competency, while supervision aims at enhancing its motivation and evaluating actual utilization of competencies.

Personnel training is often the most important part of the health development strategy, since by improving the competency of the personnel it enhances the quality of care. In planning training progammes, it is essential to define teaching or learning objectives. Evaluation analyses how the different phases progress and their quality (quality of educational objectives, of teaching, of evaluation tests). Evaluation of implementation is concerned with improvement of the students' performance.

The value of supervision is twofold. First, it completes evaluation of training by yielding information on the actual use of the competency of different types of personnel (evaluation of results). This is actually an evaluation of the quality of provision of services, based on the same criteria as those used for evaluating performance during training. Secondly, supervision enhances motivation in the personnel by listening to workers and helping them to solve the problems they encounter.

TRAINING

Training includes three phases :

- definition of the educational or teaching objectives of the training, on the basis of the tasks to be performed by the personnel (as defined in the planning phase) ;

- development of modules and of an agenda for training ;

- the training session itself and evaluation.

Teaching objectives

It is most important to define teaching objectives, since they form the basis for any evaluation. They define what the participants should be capable of doing at the end of the training session : objectives for knowledge (an intellectual approach), know-how (practical skills) and personal contact (attitudes in human relations). To be more specific, the level to be attained by the trainee is defined.

Three levels are sufficient for each type of objective. For example, an objective for knowledge will have a factual level, demanding that the trainee remember some facts such as how Aids is transmitted ; at the analysis level, the trainee should be able to interpret data (for instance, to show how childhood diarrhoea may be fatal because of dehydration, using a diagram) ; last, at the synthesis level, he/she is capable of problem-solving, an example of which is the recognition and treatment of measles.

An objective for know-how starts with imitation, the trainee performs the action in a simulated situation (recognizing the type of presentation in delivery, using a plastic head) ; at the control level,

he/she performs the action in a real situation, adjusting to specific conditions (infusing a patient). Last, at the functioning level, he/she acts rapidly and effectively, out of habit.

Objectives for contact may begin with receptiveness, in which the trainee is aware of what is happening (perceiving that a patient is worried, for instance). At the response level, he/she is capable of responding (reassuring the mother whose child has been hospitalized). Last, at the interiorization level, he/she knows how to adjust his/her attitude to the other person's value system (generating active participation and acceptance by mothers during an education for health session, for instance).

Aside from practical skills, learning is invisible. If evaluation is to be feasible, knowledge and attitudes must be translated into behaviour which is observable. The objective "knowing what anatomical changes occur in women 9 months pregnant" is not observable, whereas the objective "identify, on a drawing, those parts of the woman's womb at 9 months" is observable : the trainee is or is not able to identify those parts. The teaching objective should define what behaviour is to be observed. If the actual skill is difficult to observe, several observable acts which are indicators of performance are used.

Last, behaviour should be quantified by grading : 0 or 1 may be used, but a scale ranging from - 2 to + 2, or from 0 to 4 is more accurate. Quantitation makes it possible to set a lower limit for performance (criterion), on the basis of which to determine whether the objective has been achieved.

Let us take a concrete example. The objective is to reassure the mother whose child is hospitalized. It is quantitated using 3 indicators of performance measured by observing the nurse in role play and grading performance from + to - 2 (see table 3). The objective is reached if the total grade is 2 or more.

Evaluation of this first phase looks at each objective and makes sure that each one actually contributes to competency for a particular task (relevance) and is defined in concise, unambiguous and non-contradictory terms (accuracy). Each objective should also be observable ; that is, based on acts that are visible, measurable - that is, which may be graded, and with an acceptable level of performance - and realistic, in that it is attainable within the time allotted and with the available means.

Modules and agenda

Sessions are organized in the form of modules, each of which has objectives for teaching and contents - that is, a body of knowledge, skills and attitudes to be transmitted to trainees - and an educational method. The latter may involve questioning trainees, having them discuss the subject together, making field observations, performing technical actions in front of them, putting them in mock situations through role-playing, etc.

Whenever possible, the evaluation situation should be similar to the actual situation in which the trainee will be placed.

Table 3

Measurement of a teaching objective

Grade	- 2	- 1	0	+ 1	+ 2
1. informs mother on child's state	refuses to answer questions	does not give information	gives vague information	gives detailed information but does not answer questions	gives detailed information and answers questions
2. explains what has been done to the child	uses many medical words without explaining them	uses many medical words and sometimes explains them	uses some medical words and sometimes explains them	uses some medical words and always explains them	uses only words that the mother can understand
3. explains where to find the nurse on night duty if needed	says the night nurse should not be disturbed	refuses to respond to the mother's request	does not mention this point	just says that there is a nurse on night duty	explains where the night nurse's room is

Source : according to Guilbert.

Observation in real or simulated situations is therefore preferable, especially for practical skills and attitudes.

However, these methods are time-consuming, and require many teachers, so that evaluation is sometimes confined to written tests. Questions such as "diagnosis and treatment of malaria" should be excluded, since they are too dependent on the trainees' ability to write. Multiple choice questions (MCQ) or short, open questions (SOQ) should be preferred. Here is an example of a MCQ : which of the following situations contribute to malnutrition in infants? : (a) diarrhoea, (b) measles, (c) prolonged absence of the mother, (d) scabies, (e) pertussis (1).

In MCQ tests, the question should be simple, concise and unambiguous, negations, when unavoidable, should be underlined, the answer should not be induced, and only plausible snares (wrong answers) given.

Here is an example of a SOQ : Mrs. Lukunga, age 26, comes in to give birth to her 3rd child. Labour has been proceeding for 4 hours and dilatation is 6 cm. Although her uterine contractions are satisfactory, dilatation does not progress during the next 2 hours, and oedema develops on the cervix. What is your diagnosis? What do you do (2)?

The SOC question should be simple, concise and unambiguous, pertain to a major point that has been taught, not require an overlong response. Room should be left under the question for the answer, and the grid for correcting it prepared in advance.

---

(1) Answers : MCQ = A, B, C, E.

(2) SOQ = cervical dystocia, arrested dilatation. Spasfon, IV, 1 vial and induced rupture of membranes.

MCQ and SOQ are quite appropriate to evaluation of knowledge, but not to evaluation of know-how and attitudes, since they in fact only evaluate knowledge linked with skills and attitudes.

In the last phase, the programme for the session is designed, leaving enough time for each module. Whenever possible, it is a good idea to begin the session (or the module) by an evaluation exercise, called a pre-test. The latter gives trainers an idea of the starting level of trainees, and points to the areas requiring emphasis. Furthermore, the final evaluation or post-test will show what progress has been made between the two tests. To be comparable, pre and post-tests must be identical.

Evaluation of this second phase is based on the following criteria : are the contents of the modules written down? Are the teaching methods appropriate to the trainees and to resources? Are the evaluation tests appropriate? Is the programme for the session ready? Are trainers and trainees informed of the dates of the session?

Session and evaluation

During the training session, three evaluations are needed : of how the session functions, of the trainees' performance and of the course itself.

Evaluation of functioning is the easiest, and consists of making sure that the target population was really reached :

$$\text{Participation (\%)} = \frac{\text{no. of trainees present}}{\text{planned no. of trainees}} \times 100$$

It is also important to make sure that the programme was respected ; that is, that the modules were completed, the agenda followed and the participation active. Progress by trainees is measured by final evaluation, but pre and post-tests should be given whenever possible.

A single test yields an indicator of performance, or the percentage of trainees who are operational at the end of the course (on the basis of the criteria defined for each teaching objective) :

$$\frac{\text{number of "operational" trainees}}{\text{number of trainees tested}} \times 100$$

If pre and post tests are given, the progress achieved by each trainee may be calculated by subtracting the grade received on the pre-test from the post-test grade. The average progression within the group will be the ratio between the total progress of trainees and the number of trainees tested. Relative progress will be the ratio between progress for each trainee and maximum progress (highest grade less the grade obtained on the pre-test). The average relative progression may also be calculated using each individual's relative progress.

For example, trainee A had 12/20 on the pre-test and 16/20 at the post-test, whereas B had 10/20 and 14/20.

Progression of A = 16 - 12 = 4

$$\text{Relative progression of A} = \frac{4}{20 - 12} = 0.5 \text{ or } 50\ \%$$

Progression of B = 14 - 10 = 4

$$\text{Relative progression of B} = \frac{4}{20 - 10} = 0.4 \text{ or } 40\ \%$$

Evaluation of the course is also based on the pre and post-test results, by ventilating these for learning modules rather than for trainees. With a post-test, the average grade of trainees and the proportion of those who are operational are calculated for each module. With a pre and a post-test, the average progression of trainees and their relative progression (in %) are calculated for each module. Modules may be compared, and the most successful contents and methods determined.

The quality of the evaluation tool is also analysed, especially for MCQ and SOQ. To do so, trainees are divided into two groups : one achiever group, including the upper third with the highest grades on the post-test, and a weak group, containing the third of the class with the lowest grades on the post-test.

Two ratings are calculated for each question : a rating for difficulty and one for discrimination. Paradoxically, the rating of difficulty measures how easy the question is : it is the percentage of trainees in the two groups who gave the right answer. It ranges from 0 to 100 %, and should be somewhere between 30 and 70 % if the question is good.

$$\text{Rating for difficulty} = \frac{F + f}{N} \times 100$$

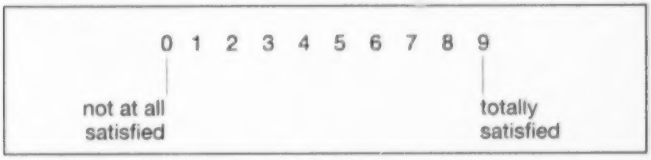
F = number of correct answers in the achiever group
f = number of correct answers in the weak group
N = total number of trainees in groups F and f

The rating for discrimination measures the extent to which the question discriminates between trainees at different levels.
It ranges from - 1 to + 1 and is at least equal to 0.25 if the question is a good one.

$$\text{Rating for discrimination} = \frac{F - f}{N} \times 2$$

Calculation of these two ratings leads to the accumulation of a list of questions for use in subsequent tests. The rating for difficulty is used to adjust the difficulty of the test to the aptitudes of trainees. An overly difficult question goes unanswered, and gives the impression that trainees have not progressed ; when the question is too easy a right answer is given at the pre-test and there is no room for advances. Rating for discrimination is less valuable in formative evaluation, since the purpose of the latter is not to select trainees.

Last, it may be interesting to ask the trainees themselves to evaluate the course. A few questions suffice : did you achieve the objectives of this session? What did you find easy or difficult? A scale may be used : for instance, are you satisfied with this session? Grade your degree of satisfaction between 0 and 9 (circle the figure chosen).



SUPERVISION

Supervision is a managerial and motivation-supportive activity comprising three phases : preparation of supervision, supervision itself and decision-making.

During preparation, documentation on the programmes (epidemiological data, objectives, programmed activities and resources, job descriptions, activities reports and evaluation reports) is collected. Next, supervision is planned (who to meet, agenda, itinerary). The criteria for evaluation speak to the different aspects of the documentation collected, the agenda established and the people informed of the visit. The minimum acceptable periodicity is once quarterly for a new activity and once a semester for a longstanding activity. The recommended periodicity is once a month for a new activity and once quarterly for an older one.

Supervision itself is a way of meeting the health teams (nurses, community health agents, health committees) and the local authorities at the grassroots level. Discussion determines how much these grassroots protagonists know about the programmes and how they accept their tasks, and identifies the reasons for any dysfunctioning or lack of motivation. Observation of the personnel at work is useful in appraising their attitudes, know-how, utilization of resources and organization of work. This identifies needs for training and follow-up. At the end of the visit, a general assessment of programmes and competencies may be made with the help of the entire health team.

The criteria for evaluation are both quantitative (number of supervisions performed or planned) and qualitative (everyone is seen individually, grids for the evaluation of competency are filled in, the meeting with the health team took place).

At the end of the meeting, decisions should be made in conjunction with the entire health team : these may deal with the identification of needs for training, changes in schedule or in the weekly working agenda, urgent logistic needs, etc. Last, all decisions and the calendar for their implementation should be noted by the supervisor in a report that is sent to the people involved and to the higher echelons.

Criteria for evaluation include the number of decisions made by general agreement, whether a person is put in charge of and an agenda established for implementation of each decision, a new working plan is put into writing and the supervisory report is written and sent to all individuals involved.

Self-evaluation exercise n° 6

In the framework of malaria control, refresher training in treating malaria in children under 5 is planned for nurses. The purpose is to assure provision of routine treatment with chloroquine and aspirin for all children with fever (38 °C or over), irrespective of the cause of the fever. Doses of chloroquine are 10 mg/kg on the first and second days and 5 mg/kg on the third day. For aspirin, the dose is 50 mg/kg/day in at least three divided doses, repeated as long as fever persists.

Many children are presently given injections of quinine, since parents view injections as the most efficient treatment. They are in fact dangerous since there is a risk of abscess or paralysis of the sciatic nerve. Quinine may be administered in infusion, but this is a difficult, costly procedure in small children. Injectable treatment should be reserved for patients suspected of pernicious episodes (temperature above 40 °C, convulsions or behavioural disorders), or for those with severe vomiting making oral intake of drugs impossible, or again, in case of resistance to chloroquine.

1. Define the teaching objectives of this training. Specify the type, the evaluation situation and the criterion.

2. How will you evaluate this training? Define the actual procedure.

Answers at the end of this issue.

EVALUATING THE HEALTH STATUS OF A POPULATION

To determine the health status of a population, a survey based on three types of indicators - mortality, morbidity and nutritional indicators - is required.

The ultimate goal of all community health programmes is the improvement of health. In a direct health action strategy, programmes directly affect the population's health, the improvement of which is the objective, or expected result. In health development and economic development strategies, the effects of health programmes on the population are indirect and are seen through evaluation of impact.

Measurement of positive health status seems to be the most valuable way of establishing a single overall gauge of the various effects of the programme. There is unfortunately no single indicator of this type for the evaluation of the results of a health programme in a small area, since they are all either too subjective or obtained from a sum of data requiring numerous measurements. "Negative" indicators of health - mortality and morbidity - are most often resorted to.

INDICATORS OF MORTALITY

Indicators of mortality yield imperfect information on health status, since they only consider the most severe complication of disease : death. And yet they are invaluable, since they reflect socioeconomic conditions and the state of the health system, and are therefore frequently used at the national and international level for intercountry comparisons. When working on district health programmes, it is tempting to use these indicators to measure the effectiveness of programmes as well.

The crude mortality rate (number of deaths per 1,000 inhabitants in a given year) depends on the age pyramid, and is therefore not very useful ; there is a tendancy to use rates for certain age groups, all causes included.

For intrauterine deaths, the late foetal mortality rate or intrauterine mortality rate (per thousand) is used :

$$\frac{\text{no. of still-births after 28 weeks of amenorrhoea}}{\text{no. of live and still-births}} \times 1{,}000$$

For infants, the infant mortality rate (IMR) (per thousand) is used :

$$\text{IMR} = \frac{\text{no. of deaths between birth and first birthday}}{\text{no. of live births}} \times 1{,}000$$

For children, the child mortality rate (CMR) (per thousand) is used :

$$\text{CMR} = \frac{\text{no. of deaths between first and fifth birthday}}{\text{no. of children between first and fifth birthday at midyear}} \times 1{,}000$$

and the under-five mortality rate (U5MR) (per thousand) :

$$U5MR = \frac{\text{no. of deaths between birth and fifth birthday}}{\text{no. of children between birth and fifth birthday at midyear}} \times 1{,}000$$

U5MR is not equal to IMR + CMR, since they have different denominators. This makes comparison between infant mortality (< 1 yr) and child mortality (1-5 yrs) difficult. To simplify, the number of live births may be used as the denominator, and an infant-child mortality rate (per thousand) defined :

$$ICMR = \frac{\text{no. of deaths between birth and fifth birthday}}{\text{no. of live births}} \times 1{,}000$$

All of these indicators are usually calculated for one year.

For maternal health, the "rate" of maternal mortality, or the maternal mortality ratio (MMR) (expressed as deaths for 100,000 births) is used :

$$MMR = \frac{\text{no. of deaths during or within 42 days of end of pregnancy}}{\text{no. of live births}} \times 100{,}000$$

Note : accidental or incidental causes are excluded. If sufficient follow-up is not feasible, the period following pregnancy may be reduced from 42 to 7 days.

Last, in programmes concentrating on one specific disease or a group of diseases (infectious diseases, for instance), the mortality rate for that disease may be used ; that is, the number of deaths caused by that disease for the population at midyear (expressed in deaths per 100,000 inhabitants).

Assay methods

Use of mortality rates requires a data-collection system in which deaths are notified to a civil registration bureau. Preconditions for validity include notification of all cases, with accurate specification of age, sex, place and cause of death. These prerequisites are rarely achieved in third world countries. Notification may be improved by the health service itself, or a survey may be made to estimate these indicators.

Surveys encounter another obstacle, however, which is the size of the sample required to determine these (relatively rare) events sufficiently accurately. To measure an infant mortality rate of 100 plus or minus 10 per thousand, there must be at least 3,000 births a year, which in turn implies a sample population of at least 50,000. For a maternal mortality rate of 1,000 per 100,000, or 10 per thousand + or - 2.5 per thousand, there must be 5,000 births a year, or a population of 125,000. These samples are too large for an evaluation team with limited resources.

There are other, less cumbersome methods for estimation : these are indirect methods. Instead of observing the event directly,

adults (> age 15) are asked how many of their sisters died during pregnancy, childbirth or immediately thereafter, or mothers are asked how many live children they brought into the world, and how many died.

With indirect methods, a person questioned once yields information on all of his/her sisters from the beginning of their reproductive years, and a mother gives information on all of her children for their entire lifetime. This multiplicatory effect makes it possible to divide the sample by 10, so that the maternal mortality rate may be determined by questioning of 3,000 adults using the sisterhood method, and the under-five mortality rate may be based on questioning of 300 mothers using Brass' surviving children method. On the other hand, indirect methods yield values corresponding to a period several years prior to the survey, so that the effects of the programme may not be very visible. Measurement of mortality by indirect methods is definitely useful, but in itself it is insufficient for evaluation of the effectiveness of a community health programme.

INDICATORS OF MORBIDITY

Since disease is far more frequent than death, morbidity is easier to measure on small samples. There is one overall indicator of morbidity and mortality : it is the number of good-health-days lost. For example, acute diarrhoea may represent 10 good-health-days lost. Severe dehydration leading to death will represent the loss of all of the days the child might have expected to live ; that is, his/her life expectancy at the time of death. Last, an ailment leaving lasting sequelae represents a loss of life expectancy at the age of partial recovery multiplied by the coefficient of incapacitation.

This indicator is in fact rarely used, except by economists who use it to calculate the cost of disease. It should be better known, since it takes the beginning of a disease, its length and its progress toward recovery, sequelae or death into consideration. When evaluating a programme directed at several high-priority diseases, good health days lost because of those diseases might be considered. Since this is not easy to calculate, two other indicators are usually used : prevalence and incidence.

Prevalence and incidence

Prevalence represents the entire diseased population at a given time (point prevalence). Period prevalence means all people affected at any time within a specified period (prevalence over a month, or year...).

$$\text{Prevalence (‰)} = \frac{\text{total number of cases}}{\text{exposed population}} \times 1{,}000$$

Incidence represents the number of people who contract the condition during the given period (monthly, annual incidence...). This shows new cases only.

$$\text{Incidence}_{(‰)} = \frac{\text{number of new cases}}{\text{exposed population}} \times 1{,}000$$

The denominator, or exposed population, represents either the entire population or an age group, one or the other sex or a geographically limited portion of the population, depending on the disease. Figure 10 shows an example of how incidence and prevalence are calculated.

In a village of 500 inhabitants, each case of measles is shown on the table by a bar, the length of which corresponds to the duration of the disease and its outcome (+ = death). For point prevalence patients are counted on 1st January ; that is, 2 cases : 500 inhab. = 4 per thousand.

For prevalence during the first quarter, everyone who was diseased at any point during the quarter is counted ; that is, 9 cases : 500 inhab. = 14 per thousand.

For children under 15 (45 % of the population), the same rates can be calculated for 225 inhabitants (500 x 0.45), provided we are sure that all patients were under 15.

Incidence and prevalence are not independent : prevalence for a given period is equal to the sum of point prevalence at the beginning of the period and incidence during that period (in the above example, 18 = 14 + 4).

Furthermore, incidence and prevalence for a given period are practically the same if the duration of illness is short in comparison to the period measured. Last, in a stable situation, one in which the disease is neither increasing nor decreasing, point prevalence is equal to incidence x average duration of the disease. The average duration of disease is expressed in units of periods of measurement of incidence. In the above example, the average duration of measles is 4 : 14, or 0.29 trimester, or 26 days.

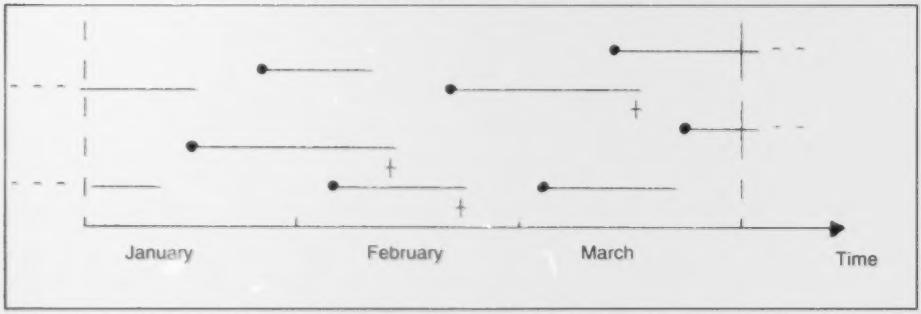


Figure 10 : Calculating prevalence and incidence.

Which indicator should be chosen, then? Prevalence or incidence? Incidence measures a recent situation, the onset of disease. It is therefore an indicator of choice for the evaluation of primary preventive action attempting to avoid disease. It is not easy to measure, however : a retrospective survey can hardly go back more than 2 weeks, since memory is not sufficiently reliable beyond that, and repeated surveys are not feasible for financial reasons.

Prevalence shows an overall situation, with individuals at different stages of disease. It reflects both onset of disease and its duration. It should be used to evaluate curative as well as secondary and tertiary preventive action attempting to reduce the duration of active disease and the human disease reservoir. For chronic diseases, point prevalence is easy to measure, by a single survey, if the condition is sufficiently widespread.

NUTRITION-RELATED INDICATORS

Nutrition-related indicators are useful for evaluating the results of routine Maternal and Child health consultations and of the distribution of nutritional supplements, as well as the impact of nutrition education. They are particularly valuable in children under five and pregnant women, who are the target populations for these programmes.

Nutrition-related indicators for pregnant women

The nutritional status of pregnant women may be evaluated using infant weight at birth ; this is also the point of departure for monitoring infant growth. The criterion is the percentage of children weighing over 2,500 grams at birth :

$$\frac{\text{number of live-born babies weighing 2,500 g or +}}{\text{total number of live births}} \times 100$$

The norm is 90 %.

Criteria for high-quality measurement are an accurate, reliable scale, proper handling of the child and of the scale (setting it at zero and reading).

Another indicator is surveillance of pregnant women's weight gain, enabling individual follow-up during pregnancy, whereas birth weight yields post-pregnancy information. Interpretation of this weight increment is delicate, but may be done with the help of a Weight/Height/Week calculator for pregnancy (cf. N 181/182 of this journal).

Nutritional indicators for children under 5

There are three ways of evaluating nutritional status in infants and children under five : weight, height and arm circumference (cf. figure 11).

The child is weighed naked, on a baby scale or, for older children, a spring scale (the child is seated in hanging panties). Care is taken to make sure the scale is accurate and set at zero, and that the child does not move about. Consistent use of the same scale is preferable.

Children under two are measured in a supine position (the term used here is "length") along a graduated measuring rod. Two

people are required : one holds the head firmly against one end of the rod, while the other holds the knees straight and presses the other end against the feet. From age two on, children are measured standing against a vertical wall, with their legs straight and their heels touching the ground and the wall (taking care to locate the zero on the ground) : a cursor is then placed on the head.

Last, measurement of arm thickness, known as arm circumference, is valuable since it is inexpensive. Arm circumference remains the same between ages 1 and 5 years, and it is easy to make a measuring tape using an old x-ray. Arm circumference is measured on the extended left arm, halfway between the olecranon (elbow) and the acromion (the edge of the shoulderblade, at the meeting of the biceps and the deltoid muscles. Care should be taken not to compress the arm with the tape.

Once these three measurements - height, weight and arm circumference - are available, four nutritional indicators may be determined :

- weight for age is widely used : it shows the child's weight for his/her age in months or years. It is an indicator of nutritional status at a given time, since weight may vary considerably with dietary intake and health status ;

- height for age shows the child's height for his/her age in months. This indicator evaluates the nutrition situation over the past two or three months ;

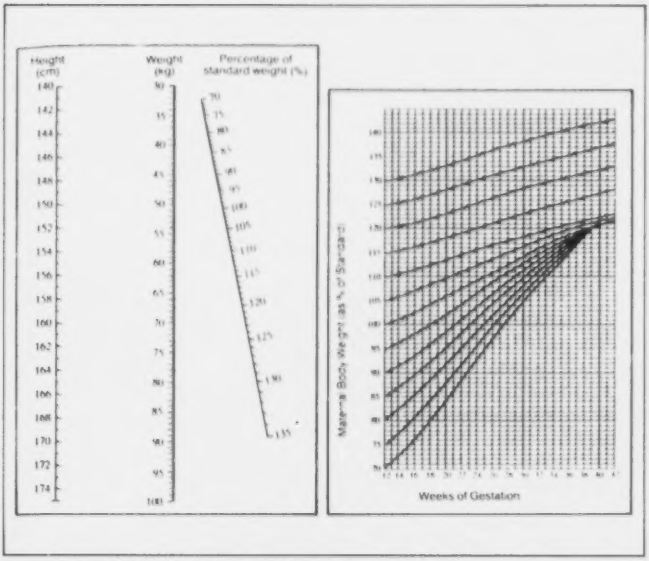


Figure 11 : Evaluation of the nutritional status of pregnant women

- weight for height studies how thin the child is. He/she is weighed, then stood up in front of a thinness chart facing a column corresponding to the weight found. The height seen determines whether or not he/she is undernourished, but this indication is only valid for acute malnutrition. It does not evidence chronic malnutrition, but has the advantage of being usable without reference to the child's age ;

- arm circumference is used directly as an indicator, since it is independent of age between ages 1 and 5.

Use of these indicators requires the definition of norms characterizing satisfactory nutritional status or malnutrition. The most commonly used references are those defined by the NCHS (National Center of Health Statistics, USA) and the WHO, when no national norms are available. A median weight and height are shown for each age and sex, along with the standard deviation and percentiles, indicating their distribution around the median.

There are three ways of determining the cut-off value separating the malnourished from the non-malnourished. First, with reference to the median. This is the easiest method : below x % of the median, the child is malnourished. However, this does not take the natural scattering of values into account. Next, the cut-off may be set at two standard deviations below the median (M - 2 SD). If the norms are available, this is easy and more accurate. Last, the cut-off may be set at the 3rd percentile : this is approximately equivalent to M - 2 SD.

In practice, the following norms are used :

- weight for age : cut-off at M - 2 SD or 3rd percentile of the national (or NCHS) norm. If impossible, cut-off at 75 or 80 % of the median ;

- height for age : cut-off at M - 2 SD or 3rd percentile of the national (or NCHS) norm. If impossible, cut-off at 90 % of the median.

For these two indicators, point prevalence of malnutrition is :

$$\text{Prevalence of malnutrition (\%)} = \frac{\text{no. of children with weight (or height)} < \text{cut-off}}{\text{no. of children weighed (or measured)}} \times 100$$

- weight for height : colours on the thinness chart correspond to the following intervals :
dark red : from 60 to 69 % of the median : severe malnutrition
red : from 70 to 79 % of the median = moderate malnutrition
yellow : from 80 to 89 % of the median = incipient malnutrition
green : from 90 to 109 % of the median = no malnutrition

The proportion of "red" children (< 80 % of the median) or the proportion of children in each colour group may be noted.

The M - 2 SD cut-off should be used whenever possible.

Arm circumference (AC) should only be used when the other indicators are not available. Malnutrition is said to be severe if AC is 12.5 or less (red) and moderate between 12.5 and 13.5 (yellow), while nutritional status is satisfactory if AC is equal to 13.5 or over (green). The proportion of children in each colour category may be noted.

WHAT IS TO BE EXPECTED OF EVALUATION?

After reading this issue and thinking about its contents in the light of his/her own experience, the reader may feel discouraged, skeptical or enthusiastic.

Discouragement may stem from the apparent difficulty of these forms of evaluation. The many computations required at the different phases and the need to implement some surveys may increase the impression of complication. The proportion of a programme's resources devoted to evaluation must not be too large. Evaluation should therefore be adjusted to the situation. Molecular evaluation and the tracer method, considering only one part of the programme, are the most appropriate methods for beginners.

Although every programme has shortcomings (since no health programme is perfect, and all must be constantly readjusted to a changing context, in order to be effective), these may be a cause for concern, when detected.

The gap between the strict logic of evaluation and the inevitable twists and turns inherent in a programme may lead to some skepticism. Management based on objectives consists of a linear succession of acts : I plan, I act, I evaluate. At all phases of the programme, however, interaction occurs between the planned programme and the system within which it functions, including the health system, the social system and the systems of production. Actually, each system reacts in accordance with its own internal logic, so that the outcome is a more or less harmonious combination.

It is important to retain a degree of distance from the outcome of evaluation, and to refrain from demanding more of the figures than they are able to tell. The present paper has attempted to indicate what is not shown by evaluation, but it does not reveal everything : it does not show the ideas and working habits acquired by people who will be able to put them into practice in other contexts years later. It does not show that a programme that was interrupted because it was not viable may still have been an enriching experience.

However imperfect, evaluation does exist, and no better method is available at the present time. Flexibility should be used in applying it, and adjusting it to circumstances : strict logic does not median rigid application.

If it evokes enthusiasm, this presentation has achieved its objective. Lucidity should prevail, however, to avoid any "technocratic" tendancies.

Two other types of evaluation have not been dealt with here. Economic evaluation may yield both economic and sociological information on the functioning of health services and on the rationality of users and officials.

Economic evaluation is concerned with action and effects. At the action level, it looks into the cost of activities, and therefore their productivity. It attempts to estimate all costs, be they budgeted (visible) or not translated into monetary terms (invisible). At the effects level, economic evaluation is interested in the cost of results, or efficiency. Cost/efficiency analysis represents the ratio between results and costs, whereas cost/benefits analysis corresponds to the ratio between impact and cost. To perform these, costs induced by the programme must be identified, be they direct or indirect. Evaluation of efficiency is the most complete evaluation of a programme.

Last, we have confined our remarks to what is known as professional, or routine evaluation ; that is, one which compares the programme and its effects to commonly accepted norms. In many cases, however, these norms are debatable or actually nonexistent. Research evaluation then aims at establishing such norms, or references for the measurement of action. The establishment of new norms and the improvement of the old ones constitutes a vast research programme.

ANSWERS TO THE SELF-EVALUATION EXERCISES

Exercise n° 1

1. Programme evaluation encompasses all of the procedures aimed at making a value judgement of a health programme by comparing action and its results with the objectives.

2. Goals are formative (improvement of the programme) and summative (making an overall assessment of the programme).

3. Control, an authoritarian process aimed solely at punishing individuals ;

- "make-believe", aimed at creating a positive image or settling scores, with no respect for truth ;

- self-flogging only views one aspect of reality, the failures ;

- follow-up does not involve a value judgement and does not cover all parts of the programme.

4. Compare your answers to the list given in the chapter.

Exercise n° 2

1. Direct health action aims at reducing mortality and morbidity by provision of curative and preventive care, health education and sanitation measures...

Health development aims at improving the health system by developing its human resources (personnel training, community-oriented action) and its material resources (drugs, equipment, facilities...) ;

Economic development aims at improving the standard of living of the population through productive activities (farming, crafts, trade...).

2. An activity is represented by the act done. An objective is the change induced by the activity ; it is a direct effect of the activity.

Example : - vaccinating, training (acts)
- immunization, competency (effects)

3. Evaluation of resources
Evaluation of activities
Evaluation of implementation

4. Evaluation of results
Evaluation of impact
Evaluation of strategy

Exercise n° 3

1. Evaluation of resources : this is not feasible, since the data is not available.

2. Evaluation of activity

• Visits planned : one a month to each village or neighbourhood, or :
- 6 x 4 = 24 visits to villages ;
- 6 x 4 = 24 visits to neighbourhoods of Obamo City.

• Visits made : 3 + 2 + 4 + 3 = 12 visits to villages, or 12 : 24 = 50 %. This is insufficient. In neighbourhoods, 22 visits, or 22 : 24 = 92 %. This is adequate.

• Meetings held :

- 10 : 24 = 42 % in villages. This is insufficient.

- 19 : 24 = 79 % in neighbourhoods. This is satisfactory.

Conclusion : Alphonse does not go to villages often enough.

3. Evaluation of implementation
Target population = (12,000 + 6,000) x 20 : 1,000 = 360 cases.

• Case-finding
total of new cases seen by Alphonse in 6 months : 21 + 7 + 15 + 8 + 10 + 12 = 73.
73 : 360 = 20 % of the target population detected in 6 months : this is satisfactory, since at this rate everyone may be detected over a 3-year period.

• Referral
total of new cases seen by the doctor in 6 months = 41.
41 : 73 = 56 %, the doctor only sees 56 % of the cases referred to him/her by the nurse : this is insufficient.

• Follow-up of treatment = treated : registered = 58 : 62 = 0.94 in January, etc.

$$\text{Mean for 6 months} = \frac{0.94 + 0.78 + 0.97 + 0.76 + 0.96 + 0.89}{6} = 0.88$$

On the average, 88 % of patients were treated each month. This is good, especially considering how many times the villages were not visited.
Conclusion : too many detected patients do not go to see the doctor.

Exercise n° 4

UTILITY : what effect did the action have on the people who participated in the programme?
The effects should be determined for both aspects of action : education for health and nutritional rehabilitation.

For prevention, the children's old growth charts may be reexamined to determine what percentage was undernourished. For children aged 1-2 years, three years ago (before the action) and now. Make sure the same periods of the year are compared, so that seasonal variations will not distort comparisons.

For rehabilitation, old records are reexamined for comparison of children when discharged from the centre with their nutritional status when entering it. It would be a good idea to have a measurement taken several months after leaving the centre.

EFFECTIVENESS : what was the effect of the action for the population of the area as a whole?
The effectiveness of this programme should be seen in a lowered percentage of undernourished children in the target population.

This percentage should therefore be measured, using a survey of a representative sample of the target population of the area. But what will be the other term of the comparison? No measurement prior to the implementation of the programme is available here. It is best to compare figures with a neighbouring area to which the programme is to be extended (in addition, this will constitute a prior measurement for that area). Care must be taken to make sure that the two areas are comparable with respect to major factors such as dietary habits, age structure, etc.

IMPACT : what were the overall effects of the programme?

Questions to be asked may include : did children's diet change? Was there a decline in infections for the age group touched? Did leading women introduce changes in the organization of village women? Is childhood malnutrition perceived differently now by health workers? etc.
A survey of a representative sample of children in the area will yield a response to the first two questions (here too, there is the question of a reference measurement). For the others, talks with villagers, women leaders, administrative officials or traditional notabilities and health workers will suffice.

Exercise n° 5

1. Computation of the target population
Expected annual number of pregnancies = total population x birth rate = 38,000 x 40 : 1,000 = 1,520

• Availability (time) : over the year, 6 individuals for 6 hours a week, or 6 X 6 X 52 = 1,872 hours during which service is actually available.
Needed availability : 1/2 h + 3 x 1/4 h for each woman, for the 4 ANC, or 1.25 h. Thus, 1,520 x 1.25 = 1,900 hours.
Rate of availability = 1,872 : 1,900 = 99 %

• Physical accessibility
Population living less than 5 km from a health centre : 20,000 x 0.8 + 5,000 x 0.7 + 8,000 x 0.5 + 4,000 x 0.5 + 1,000 x 0.9 = 26,400.

$$\text{Accessibility} = \frac{\text{population} < 5 \text{ km distant}}{\text{total population}} = \frac{26{,}400}{38{,}000} = 69\ \%$$

Accessibility x availability = 0.69 x 0.99 = 69 %

• Utilization = no. of record cards for 6 months/expected pregnancies over 6 months
Expected pregnancies over 6 months = 1,520 : 2 = 760
Utilization = 575 : 760 = 76 %

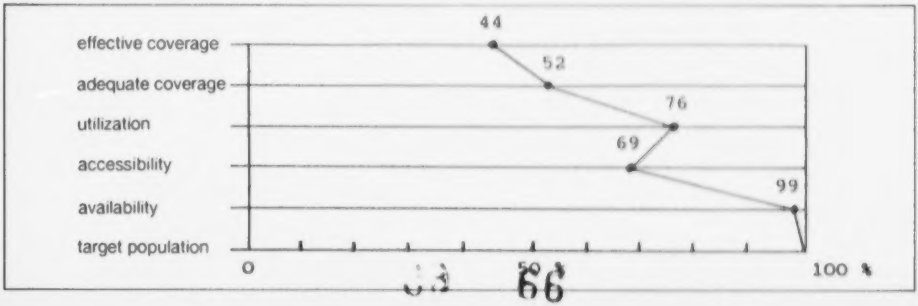
• Adequate coverage = no. of women > 2 ANC over 6 months/expected pregnancies over 6 months = 393 : 760 = 52 %

• Effective coverage = adequate coverage x percentage of high-quality ANC
Percentage of high-quality ANC = 1 - (incomplete cards : total cards) = 1 - 91 : 575 = 0.84 or 84 %.

Note : it would be preferable to dispose of data pertaining to the quality of the visits rather than to record cards (which may cover one or several visits).

Effective coverage = 0.52 x 0.84 = 0.44 or 44 %

2. Coverage diagram



3. Conclusions

First problem : accessibility. Can ANC be brought closer to the population by training community health agents, running mobile consultations or creating new centres?
Second problem : women do not come back for the last one or two ANC. Why? Should education for health address this subject?

Exercise n° 6

There are several possible formulations. The answers given here are not the only acceptable ones.

1. Teaching objectives : at the end of training, nurses should be able to :

a) list the reasons for routine treatment of fever by chloroquine and aspirin. This is an objective for knowledge. Evaluation : question the nurse on the reasons for this treatment. Criterion : he/she mentions three reasons : control of malaria, danger and cost of injectable treatment ;

b) list exceptions to routine treatment. This is an objective for knowledge. Evaluation : question the nurse on exceptions to treatment. Criterion : lists at least 4 of the 5 reasons ;

c) prescribe the amount of drugs corresponding to the patient's weight. This is an objective for knowledge. Evaluation : give the nurse weights for 5 children ; he/she must determine the prescription. Criterion : right doses for at least 3 out of 5 cases, and all within the acceptable limits (+ or - 25 %) ;

d) convince parents of the need for oral treatment. This is an objective for human relations. Evaluation : simulated situation (role-playing) : another nurse plays the role of the child's parent. Criterion ≥ 2/4.

0 : refuses to explain the reasons for the treatment.

1 : just says that it is a new compulsory treatment plan.

2 : explains the dangers of injections.

3 : idem 2 and explains that it is less expensive and highly effective.

4. idem 3 and explains that the fever is not serious and that the parents may come again if necessary.

Note : the ability to measure the child's temperature and weight and to calculate the required dose are assumed to be prerequisites for the course.

2. Evaluation of the result : it is not enough to evaluate achievements at the end of training : there is the question of whether what has been learned is used in everyday practice. Evaluation of results is done in the course of supervision :

- by questioning the nurse to see whether he/she still remembers objectives a and b :

- by observing the nurse during the consultation to see whether objectives c and d are achieved ;

- by examining the consultation record book to determine the proportion of children under 5 with fever who received the standard treatment and the proportion of those who were given another treatment for justifiable reasons. These percentages may be followed in time to determine whether they change significantly (ideally, the sum of the two should reach 100 %).

Criteria for evaluation of objectives a, b and c during training may also be used for evaluation during supervision.

FURTHER READING...

General publications on evaluation

INSTITUT DE SANTE ET DEVELOPPEMENT. L'évaluation des actions de santé dans les pays en développement. Paris : ISD, 1988. 125 p.

MAC MAHON R, BARTON E, PIOT M. On Being in Charge. A guide for middle-level management in primary health care. Geneva : WHO, 1980. 366 p.

PINEAULT R, DAVELUIY C. La planification de la santé, concepts, méthodes, stratégies. Montréal : Agence d'Arc Inc., 1986. 480 p.

ROUGEMONT A, BRUNET-JAILLY J. Planifier, gérer, évaluer la santé en pays tropicaux. Paris : Doin, 1989. 751 p.

WHO. Health Programme Evaluation. Guiding principles for its application in the managerial process for national health development. Geneva : WHO, "Health for All" Series, 1981 ; 6. 47 p.

Criteria and indicators for evaluation

CHAULIAC M, MASSE-RAIMBAULT A.M. Nutritional Status : the interpretation of indicators. Children in the Tropics. Paris : International Children's Centre, 1989 ; 181/182. 81 p.

GUILBERT J.J. Educational Handbook for Health Personnel. Geneva : WHO, Offset publication, 1981 ; 35. 350 p.

KNIPPENBERG R, LEVY-BRUHL D et al. The Bamako Initiative : Primary Health Care Experience. Children in the Tropics. Paris : International Children's Centre, 1990 ; 184/185. 94 p.

WHO. Development of indicators for monitoring progress towards health for all by the year 2000. Geneva : WHO, "Health for All" series, 1981 ; 4. 91 p.

WHO. Indicateurs pour la surveillance des progrès réalisés vers la Santé pour tous en l'an 2000 au niveau du district. Rapport d'un groupe de travail. Brazzaville, 1989. 37 p. et annexes.

Surveys and statistics

BRASS W. Methods for estimating fertility and mortality from limited and defective data. Carolina : Chappell Hill, 1975 : 50-59.

FILIPPI V, GRAHAM W. Estimer la mortalité maternelle à l'aide de la méthode des soeurs. Paris : CEPED. Dossier du CEPED, 1990 ; 13. 30 p.

GRAHAM W, BRASS W et al. Indirect estimation of maternal mortality : the sisterhood method. London : Center for Population Studies Research, 1988. 30 p.

LUTZ W. Santé de la collectivité. Rennes : ENSP, 1983 ; six volumes.

RUMEAU-ROUQUETTE C, BREART G, PADIEU R. Méthodes en épidémiologie. Paris : Flammarion, 1985. 398 p.

VAUGHAN JP, MORROW RH. Manual of epidemiology for district health management. Geneva : WHO, 1989. 198 p.

AVAILABLE ISSUES

■ 1982

138/139/140** How to feed young children

■ 1983

146 Third world women at work

■ 1984

147/148* The principal foods : their composition, conservation and transformation
149/150* The ongoing evaluation of protein-energy malnutrition
152/153* Prevention in obstetrics

■ 1985

154/155* Education on sexuality, the family and society today
156/157* Primary rehabilitation care in Mali
159 Women's lives, mothers' health

■ 1986

160/161* From routine... to epidemiology... to health activities
162/163* Immunization
164 Early mother - Child interaction
165 Xerophthalmia andblindness of nutritional origin in the third world

■ 1987

166 Teaching nutrition to young children
167/168* Weaning foods
169 Major haemoglobinopathies
170/171* Proceed with caution... Children under 6

■ 1988

172 AIDS, the mother, the child
175/176* Strategies for combating endemic goitre
177 Diet, environment and children's development

■ 1989

178 Malaria
179 Genetics and health
180 Health for adolescents and youth
181/182* Nutritional status : the interpretation of indicators
183 Learning how to read... Why ?

■ 1990

184/185* The Bamako Initiative : primary health care experience
186 Nutritional anaemia
187/188* Childbearing and Women's health
189 Immunity and Nutrition

■ 1991

190/191* New Borns... getting a good start
192 Nutrition education
193/194* Controlling fertility
195 The developing child : tools for monitoring

■ 1992

196/197* Childhood tuberculosis, still with us...

*Double issue.
** Triple issue.

TO BE PUBLISHED

- Food and its small-scale processing
- Anthropology and programmes for development
- Food and diet at different ages

SUBSCRIPTION OR ORDER FORM

LAST NAME : .....

FIRST NAME : .....

PROFESSION : .....

ADRESS : .....
(Street or P.O. Box)

TOWN : ..... POST CODE : .....

COUNTRY : .....

I order issues numbers : .....

.....

..... simple issue(s) x US$10 (50FF) ☐

..... double or triple issue(s) x US$14 (75 FF) ☐

I subscribe (1) to the journal for US$40 (220FF) ☐

ENCLOSED FIN THE TOTAL AMOUNT OF ☐

by bank transfert to Société Générale ☐ *
RIB 30003 03300 000200091 42 70

by check (FF) ☐ *

by money ☐ *

to the order of CENTRE INTERNATIONAL DE L'ENFANCE

RETURN TO : L'ENFANT EN MILIEU TROPICAL
Centre International de l'Enfance
Château de Longchamp
Bois de Boulogne
75016 PARIS (FRANCE)

(1) 6 issues per year

NB : A special price could be allowed for an important order.

Founder : Prof. J. SÉNÉCAL
Managing Director : Prof. H. DUPIN
Editor in Chief : Dr. A.-M. MASSE-RAIMBAULT
Imprimerie Rosay, 47, avenue de Paris, 94300 Vincennes - Dépôt légal : 3e trimestre 1992 - N° 19260
N° de commission paritaire : AD 1622 - N° ISSN : 0379-2269

EVALUATING HEALTH ACTION IN THE THIRD WORLD

1992 - N° 198

After several years in the Congo, where he was in charge of health activities and specialized in public health, the author wishes to share his knowledge and experience in order to acquaint readers with one phase of action that is often particularly difficult to implement fully : evaluation. The present document uses simple, comprehensible terms, illustrated by a number of examples and followed by self-evaluation exercises, to present the different sectors and phases of health and training programmes and the steps in their evaluation, in an attempt to indicate how dynamic and enriching evaluation may be, and how necessary. It also shows that when this approach is integrated in programmes it serves as a springboard for the readjustment and possible completion of certain tasks, phases or orientations. This attempt at demystification will perhaps show many health officials that they have been doing evaluation without knowing it.

With thanks to Micheline d'Agostino - ICC for the cover illustration

INTERNATIONAL CHILDREN'S CENTRE
Château de Longchamp
75016 Paris - France
Tél. : (1) 45 20 79 92 • Télex CIENFAN 648 379 F
Fax : (1) 45 25 73 67

BEST COPY AVAILABLE